Photogrammetric shape reconstruction of diffuse and specular objects in time

DISSERTATION

zur Erlangung des akademischen Grades eines

**Doktors der technischen Wissenschaften
(Dr. techn.)**

unter der Leitung von

Univ.-Prof. Dipl.-Ing. Dr. Josef Jansa

eingereicht an der Technischen Universität Wien

Department für Geodäsie und Geoinformation

Forschungsgruppe Photogrammetrie

E120.7

von

mgr inż. Ewelina Rupnik

Matrikelnummer 1028557

11 rue Hildegard von Bingen, 94120 Paris

Wien, 15.09.2015

_____

mgr inż. Ewelina Rupnik

Photogrammetric shape reconstruction of diffuse and specular objects in time

DISSERTATION

submitted in partial fulfillment of the requirements for the degree of

Doktor der technischen Wissenschaften
(Dr. techn.)

under the supervision of

Univ.-Prof. Dipl.-Ing. Dr. Josef Jansa

at the Vienna University of Technology
Department of Geodesy and Geoinformation
Research Group Photogrammetry

by

mgr inż. Ewelina Rupnik

Registration Number 1028557

The dissertation has been reviewed by:

_____
Univ.-Prof. Dipl.-Ing. Dr. Josef
Jansa

_____
Dr. Fabio Remondino

Wien, 15.09.2015

_____
mgr inż. Ewelina Rupnik

Erklärung zur Verfassung der Arbeit

mgr inż. Ewelina Rupnik
11 rue Hildegard von Bingen, 94120 Paris

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

_____   _____

(Ort, Datum)   (Unterschrift Verfasser)

Acknowledgements

This work finalizes a chapter of my life which contributed not only to my scientific, but also my personal development. I would like to acknowledge people whom I came across in this period, who helped me and inspired me.

In the first place, I would like to thank Josef Jansa and Norbert Pfeifer for supervising my research, providing me with constructive criticism, and for the strong support of my scientific career throughout the last 5 years. Josef Jansa, again, and Fabio Remondino, thank you for reviewing this dissertation with your sharp eyes, and in time. Your comments substantially improved the quality of this work. Further gratitude to be expressed to Fabio to have allowed me to work in his research team for the last 2 years, and for boosting my publication output.

I am also thankful to Markus Haider and Franz Rammerstorfer from the *Doctoral College on Energy Systems* for the 3 years of leadership of our Working Group, and for the numerous talks on how photogrammetry can serve their applications.

Many thanks to Camillo Ressl to have introduced me to *Orient*, and to have shared his extensive knowledge on photogrammetry with me. Photogrammetry told by you sounds like a beautiful story. Sajid Ghuffar, my friend and colleague, thank you for reflecting with me on photogrammetry and other topics, for the paper reading group initiative, and the late-night feeding me with fruits in Gusshaustrasse. Michael Hornáček, you were an important actor in the way I comprehended research, how my research interests evolved, and a person who stimulated my C++ skills – I am grateful for that. Andreas Roncat and Willi Karel, thank you for your hints and discussions during my stay at the Department.

Marcela Doubkova, Elin Högström and Felix Greifeneder, although quite far from me from the scientific viewpoint, you have been all very good friends of mine and our conversations on the surrounding reality made the working environment even friendlier and well attended. It has had the implicit impact on my research output too.

Francesco Nex and Isabella Toschi, my colleagues from the italian office, thank you for your friendship, encouragement and to have put up with me — trying to come into terms with work, PhD writing and my personal life. Edward Beamer and Caroline Frank, my two precious friends, I highly appreciate your time to proof read this thesis, and your irreplaceable company while in Vienna.

Last but not least, I would like to thank my parents to have always given me the freedom of choice as to what to do, where to go, and for believing in me in moments, when I did not believe in myself.

Abstract

A multi-view imaging system is built and equipped with a set of software tools to perform 4D (3D and time) precision measurements of diffuse and specular surfaces. It consists of three consumer grade dSLR cameras, acquiring images in single frame and video modes. The feasibility of the system is proved in a series of measurements of a 4x4m platform, floating on the water surface. Both the motion of the platform's building blocks as well as the modelling of the instantaneous shape of the water surface are put in focus.

The first part of the dissertation deals with the camera orientation and triangulation tasks. Prior to the measurements, the cameras are calibrated, that is their interior orientation, lens distortions and relative orientation are retrieved. The simultaneous retrieval of all the parameters is carried out in a self-adjusting bundle adjustment using the method of *moved reference bar*. During the measurement, the three cameras have remained static with respect to one another, thus their relative orientation is regarded constant. However, at times the position of the whole system has undergone a motion. It has been due to the observed object drift, requiring a corresponding motion of the imaging system to maintain the same field of view. This motion has been unknown yet resolved by introducing a 3D rigid transformation into the bundle adjustment equations at appropriate time instances. It is possible with a minimum of three static object points observed. The dynamic referencing and triangulation of the platform targets are simultaneously solved in a bundle adjustment routine.

The surveyed object has been signalised with circular retro-reflective targets. Thanks to their particular micro-structure, a ray of light cast onto its surface returns a strong signal back to the camera. This type of targetting is attractive due to its better image observation precision and the fact that it is less susceptible to ambient lighting, such as numerous reflections from water or other shiny parts. Nonetheless, the workplace constraints, the dynamism of the floating platform and the lower quality sensors have all contributed to occasional and temporary losses of the tracked object points. To overcome the timely gaps in the target's signal, its motion is modelled in image space with up to second derivative with respect to the position, and embedded inside the Kalman filter. The Kalman filter is adopted as a predicting tool, thereby, when anomalous target's behaviour is detected, the physical measurement is replaced with the measurement suggested by the state of the model, or simply discarded.

The second part of the dissertation deals with modelling the water surface in time. Two methods are proposed, treating the water as a diffuse and specular surface, respec-

tively. In either case the water is presumed to take the shape of a travelling sine wave reducing the task of the 3D reconstruction to resolving the wave parameters.

The first conceived method performs the modelling part purely in 3D space. Having triangulated the points in a separate and unrelated phase, a sine wave is fitted into the data in a least squares manner. The second method presents a more complete approach for the entire calculation workflow begins in image space. The water is perceived as a specular surface, and the travelling specularities are the only observations visible to the cameras. The travelling highlights implicitly carry information about the underlying shape of the surface. From the law of reflection it is known that such virtual points can be exclusively single image observations, hence the depth ambiguity must be removed with additional constraints. Besides the constraint that all points lie on the parametric surface, the incident camera ray and the exitant ray from the highlight to its real position must be symmetric and coplanar with respect to the normal at the virtual points. The latter two compose the law of reflection. The condition and constraint equations land in a single system of equations that is solved with the method of least squares adjustment.

All devised approaches are validated against data of superior temporal and spatial resolution. In parallel to the developed system, the platform has been observed by an online motion capture system (MCS). The outcome platform motions are compared in terms of *ship motions*. As for the modelled shape of the water surface, a check against the data coming from a Capacitive Level Sensor, and a check on physical targets floating on the surface is performed. The wave parameters are confronted between the two developed methods too. In all cases the outcomes agree to a high degree.

Kurzfassung

In Zentrum der Arbeit steht die Entwicklung eines Mehrbild-Aufnahmesystems und der dazu gehörenden Software-Werkzeuge, um 4D- (3D- und Zeit-) Präzisionsmessungen von diffusen und spiegelnden Oberflächen durchführen zu können. Das System besteht aus drei digitalen Amateur-Spiegelreflex-Kameras, die Einzelbilder und Videosequenzen erfassen. Die Realisierbarkeit des Systems wird durch eine Reihe von Messungen einer auf einer Wasseroberfläche schwimmenden 4 m x 4 m Plattform bewiesen. Die Dauer der einzelnen Datenerfassungen blieb unter einer Minute. Sowohl die Bewegungen der Bauteile der Plattform, als auch die Modellierung der Form der Wasseroberfläche wird in den Mittelpunkt der Untersuchungen gestellt.

Der erste Teil befasst sich mit dem Kamera-Arrangement und den Aufgaben der Triangulation. Vor den Messungen erfolgt die Kamera-Orientierung, das heißt die Bestimmung der Linsenverzeichnung, der inneren und der relativen Orientierung. Die gleichzeitige Bestimmung aller Parameter erfolgt in einer Bündelblockausgleichung unter Verwendung eines bewegten Referenzbalkens (*moved reference bar*). Während der Messung blieben die drei Kameras zueinander unverändert, sodass man die relative Orientierung als konstant betrachten kann. Gelegentlich musste das gesamte Aufnahmesystem in seiner Position verändert werden. Der Grund lag in der Veränderung der Lage des beobachteten Objektes, sodass eine Nachführung notwendig war, um das Objekt im Aufnahmeblickfeld zu behalten. Diese Positionsänderung war vorerst unbekannt, wird aber durch die Einführung einer starren 3D-Transformation in den Bündelblockausgleich zum geeigneten Zeitpunkt gelöst. Mindestens drei statische Objektpunkte müssen dafür beobachtet werden. Die dynamische Referenzierung und die Triangulation der Ziele auf der Plattform werden simultan in einer Bündelblockausgleichung gelöst.

Das untersuchte Objekt wurde mit kreisförmigen, retro-reflektierenden Zielmarken signalisiert. Durch die Beleuchtung mit Scheinwerfer warfen die Ziele ein starkes Signal zurück zu den Kameras. Diese Art der Zielsignalisierung ist von Vorteil, weil eine bessere Beobachtungsgenauigkeit gegeben ist und weil sie weniger anfällig auf Umgebungslicht reagieren, zum Beispiel ist eine bessere Unterscheidung von Reflexionen an der Wasseroberfläche oder an glänzenden Bauteilen möglich. Trotzdem haben die Beschränkung des Arbeitsplatzes, die Dynamik der schwimmenden Plattform und die begrenzte Qualität der Sensoren zu einem gelegentlichen und vorübergehenden Verlust der verfolgten Objektpunkte geführt. Um die Lücken des Zielsignalpfades zu überbrücken, werden die Bewegungen bis zur zweiten Ableitung der Position im Bildraum nachempfunden und in

ein Kalman-Filter eingebettet. Das Kalman-Filter wurde als Analysewerkzeug übernommen und immer dann, wenn ein abweichendes Zielverhalten festgestellt wird, wird die physikalische Messung durch die Messung, die vom Modell vorgeschlagen wird, ersetzt oder einfach eliminiert.

Der zweite Teil befasst sich mit der zeitnahen Modellierung der Wasseroberfläche. Es werden zwei Verfahren vorgeschlagen, die das Wasser als diffuse beziehungsweise als spiegelnde Oberfläche behandeln. In beiden Fällen wird angenommen, dass die Wasseroberfläche die Form einer fortschreitenden Sinuswelle annimmt, wodurch die Aufgabe der 3D-Rekonstruktion vereinfacht wird und zur Lösung der Wellenparameter führt.

Das erste konzipierte Verfahren führt die Modellierung ausschließlich im 3D-Raum aus. Nachdem die Punkte in separaten und unabhängigen Phasen trianguliert worden waren, wird eine Sinuskurve nach der Methode der kleinsten Quadrate an die Daten angepasst. Das zweite Verfahren stellt einen umfassenderen Ansatz für den gesamten Berechnungsprozess dar, da dieser bereits im Bildraum beginnt. Die Wasseroberfläche wird als spiegelnd angesehen, und einzig die sich bewegenden Spiegelungen können von den Kameras beobachtet werden. Sie enthalten implizit die Information über die zugrunde liegende Form der Oberfläche. Durch das Reflexionsgesetz ist bekannt, dass solche virtuelle Punkte ausschließlich Einzelbeobachtungen sein können, die Mehrdeutigkeit der Tiefe muss daher durch zusätzliche Beschränkungen ausgeschlossen werden. Außer der Einschränkung, dass alle Punkte auf der parametrischen Oberfläche liegen müssen, müssen der einfallende Kamerastrahl und der Strahl zwischen den Reflexionspunkten und deren realen Positionen symmetrisch und koplanar in Bezug auf die Normalen in den virtuellen Punkten sein. Die beiden letzten Eigenschaften werden durch das Reflexionsgesetz vorgegeben. Die Bedingungsgleichungen und die Nebenbedingungen gehen in ein Gleichungssystem ein, welches mit Hilfe der Methode der kleinsten Quadrate gelöst wird.

Alle entwickelten Ansätze werden mit Daten von höchster zeitlicher und räumlicher Auflösung validiert. Parallel zu dem entwickelten System wurde die Plattform durch ein Online-Bewegungserfassungssystem (Motion Capture System (MCS)) beobachtet. Das Ergebnis der Plattformbewegungen wird in Bezug auf Schiffsbewegungen (*ship motions*) verglichen . Was die modellierte Form der Wasseroberfläche betrifft, können die Ergebnisse mit den Daten aus kapazitativen Pegelsensoren und mit auf dem Wasser schwimmenden Zielen überprüft werden. Was die Werte für die Wellenparameter betrifft, so werden ausschließlich die Ergebnisse aus den beiden erwähnten Verfahren mit einander verglichen. In allen Fällen ist eine Übereinstimmung in hohem Maß gegeben.

Contents

x

1

Introduction

1.1 Motivation and objectives

Close range photogrammetry, or image engineering for custom-made solutions has undergone significant development in recent years. Because the sensors do not contact the measured object and operate in a rapid fashion for a desired period and at a desired scale, photogrammetric practices find enthusiasts across many application fields. Thanks to redundant acquisitions, the applied methods are backed by precision estimates and reliability measures, becoming a target when quantitative evaluation is of interest. Today, many common engineering and industrial applications can be tackled using photogrammetry with commercial tools. The workflow from image acquisition through orientation to coordinate extraction and its representation in a desired form has reached a level of full automation. Nevertheless, on certain occasions, custom solutions and expert knowledge are still sought. Individual approaches are often imposed by industrial research, where less conventional problems must be addressed.

In this dissertation, photogrammetry was evaluated as an experimental method in industrial research and development (R&D), aiming at pre-industrializing a novel solar power plant (SPP) model. The concept is novel because it envisions offshore placement. The motivation for an offshore solution stems from the fact that Europe lacks land adequate for significantly large contributions of solar power. Also, occlude-less space enhances power generation efficiency, while the seawater can be exploited as a cooling system. The advantages, however, are counterbalanced by challenges, unique to the intended offshore arrangement design features. The power generation mechanism shall adopt the so called Concentrated Solar Power, in the form of large numbers of arrays of parabolic trough collectors. For the technology to be plausible, and to make the best use of it, highly precise positioning systems and stability are required. The carrying platform must be insensitive to water wave, wind and other external perturbations. Alongside that, the goal was to construct a lightweight carrying structure, made of naturally abundant materials.

The development was divided into numerical and experimental modelling. By building a simplified model of reality in a computer environment, theoretical models simulate and thereby allow for prediction of the system's performance, subject to various testing conditions. The role of numerical modelling is therefore crucial, considering the costs, labour and time that would otherwise be necessary to build the actual model. But simulation-driven designs may be insufficient for robust system characterization, especially in the presence of unmodelled system non-linearities [Magnevall 2011]. In such cases, development of experimental methods for collecting real data is integrated into the framework. If a discrepancy exists between results, the design hypotheses must be reconsidered. The real data was collected by a photogrammetric system conceived within the scope of the dissertation.

The objective of the work was twofold; to observe the motion of the simplified version of the SPP carrying platform, and simultaneously provide information on the forces generating the motion — or to be more precise — on the shape of the water waves. Millimetre accuracy and reliability were expected. The challenges were split between the uncommon surface characteristics, and the workplace constraints. The ship testing facility was in large part occupied by the model basin and offered little space around the measurement volume. It necessitated that the shape and motion be sensed from a distance.

The carrying platform was covered with highly reflective and textureless coating, while the water rendered a mirror-like appearance. Intensity information and texture play essential roles in passive vision metrology. Without it, 3D reconstruction from images is doomed to fail. Handling difficult surfaces, i.e. not reflecting equally in all directions, with remote sensing methods remains an open question. The state-of-the-art approaches faced with materials of this genre resort to rigorous modelling of the media or convert it to an opaque appearance with the help of artificial targetting, sprays and powders. Active techniques, such as structured light or laser scanning, need no intensity gradient for smooth functioning, but fail at non-Lambertian reflections. At the same time, the sequential acquisition of the laser renders it unsuitable for observing timely phenomena.

Given all the impediments and motivation, three passive dSLR cameras were acquired for the job. To keep the instrument expenses at a reasonable level, it was decided that, rather than adopting costly imaging equipment dedicated for vision tasks, consumer grade cameras of good credibility were employed.

1.2   Contributions

The primary contributions of this work were in (i) building an imaging system, (ii) accompanying it with practices and tools to solve diffuse and specular body motion, and (iii) testing the performance of the system in operation.

Hardware   The imaging equipment employed cameras not inherently made for metrology but for pleasure. At that time unprecedented on the camera market was to bridge

the gap between video recorders, and dSLR cameras by upgrading the latter to record image sequences. This thesis sets out to assess and prove the usefulness of the newly introduced functions in the field of photogrammetry.

The devised system consisted of three consumer grade cameras, each one coupled with a high power continuous illumination source. The cameras were interconnected with one another, and with a PC so as to trigger stop or start a video acquisition. Coarse synchronisation was realised with the hand of a commercial software while fine-alignment was possible thanks to a laser dot visible in all camera views. Restricted control over the camera hardware was regarded as a serious shortcoming. The line of developments and the ultimate stand of the algorithms was the outcome of hardware and workplace stipulations, including the peculiar surface attributes of the objects of interest.

Algorithms The processing roadmap in general followed the conventional sequence of camera calibration, camera system orientation, 2D image measurements, 3D triangulation and modelling. Because the application was unconventional, at each processing step necessary amendments were proposed, implemented and evaluated. The first three steps were common for both the measurement of the diffuse and specular body motions. Triangulation on diffuse materials is collinear so it proceeds the normal way, yet on specular materials it is piecewise linear and requires special treatments.

The built three-ocular imaging system was fixed on a mobile bridge that ran across the basin. As such, individual cameras were regarded stable with respect to one another but possibly subject to a unique 3D rigid transformation. The camera calibration parameters (interior orientation and additional distortion parameters) as well as the relative orientation were found once, prior to the measurements, with the *moved reference bar* method. The reference bar effortlessly generated the spatial field of a desired shape on the spot, thereby, the entire calibration procedure happened with no interference on the workplace environment.

When the bridge underwent a motion, the constancy of the relative orientation was violated. Unless the motion had been modelled within the processing pipeline — in this thesis referred to as *dynamic referencing* — the triangulation results would have been wrong. To model that, a few points had been put in proximity of the water and their positions in images were monitored. Given a change of their image coordinates, a sequential 3D-3D rigid transformation took place. It related the reference coordinate system before the motion had occurred, with that of after the motion.

Tracking points in images, be it for calibration purposes or for a specific measurement, was often hindered by occlusions or other disturbing signals in the object space (e.g. caused by glitter). The points got lost and the measurement became corrupted. This predicament was overcome by modelling the motion of a point in image space with a Kalman Filter. If a point behaved abnormally, i.e. abruptly changing its position, velocity or acceleration, it was discarded from further analyses. Alternatively one can program other decisions to be taken by the filter, e.g. to carry on the measurement with the interpolated position.

3

As far as the modelling part is concerned, two techniques were conceived, treating the water as diffuse (*method 1*) or specularly reflecting (*method 2*). *Method 1* exploited the so-called *real* points, while *method 2* worked with *virtual* points. The points are real if its projections in images come directly from the points (e.g. physical targets), or virtual (e.g. specular highlights) if the camera sees merely the reflection of a real point. Virtual points are always single image observations, as their positions in space depend on the shape of the surface it is reflected from and view angle of the camera (Vide law of reflection). Prior knowledge of the shape of the water must be thus communicated to the bundle adjustment or else erroneous results come out.

In either of the methods, the water was modelled as a special surface that travelled in time as a 3D sine wave. The estimation problem was hence to find the wave parameters — amplitude, wavelength, period, phase — that would conform with some observations. Given clusters of the 3D triangulated real points, *method 1* fitted the travelling sine wave in the 3D data. Still, the 3D coordinates were considered independent of the 2D image observations within the fitting. *Method 2* was more rigorous in this respect. The modelling was incorporated in the triangulation phase and thanks to this link (i) the 3D coordinates and wave parameters were estimated simultaneously, consequently (ii) rigorous error propagation took place. A further advantage of the combined approach was the possibility to include single image observation into the pipeline.

1.3   Outline of the thesis

This dissertation presents a complete pipeline for the exploitation of consumer grade cameras to measure shape of diffuse and specular objects in time, at close range. Chapter 2 is aimed at non-expert readers by reviewing the fundamentals and state-of-the-art in 3D shape reconstruction. Chapter 3 reviews parameter estimation and gives insights into the most essential mathematical tool in image processing, i.e. the least squares adjustment. Recursive estimation — a variation of least squares — is also explained. Throughout the theoretical chapters, it was attempted to stress the differences in approaches common among photogrammetry and computer vision communities.

The subsequent chapters reveal the experimental part of the thesis. Chapter 4 elaborates on (i) the adopted equipment, (ii) performed calibrations of the internal and external camera parameters (when cameras are static or dynamic), (iii) 2D tracking and 3D triangulation in image sequences, plus (iv) evaluates the performance of the system with independently acquired data. This chapter deals solely with diffuse-like surfaces.

Chapter 5 describes and demonstrates the modelling of water surface shape. Two methods are presented and both rely on prior knowledge of the surface shape. *Method 1* exclusively involved modelling and used triangulated points floating on the water surface as input. *Method 2* used the triangulation and modelling in a combined adjustment. It can be thought of as a *calibrated environment* method. Physical floaters were unnecessary because the shape was retrieved solely from specular reflections of known and static points in the object scene.

Finally, in Chapter 6 conclusions are drawn, lessons learnt are disclosed and some further research ideas are given.

Shape reconstruction

In this chapter, approaches to image-based shape reconstruction will be given. Because interpretation of surface shape from brightness data from images depends on

- surface properties,

- imaging geometry,

- illumination condition,

the discussion starts with a brief introduction to the models of light transport on surfaces and through volumes, which follows with the camera geometry. Next, attention is drawn to reconstruction methods that (a) are apt for diffuse and non-diffuse surface materials, (b) use single and multi-views, (c) are commonly used within photogrammetric and computer vision communities.

2.1 Light transport

Two approaches can be used for the study of light, taking into account either its wave (physical) or particle (geometrical) nature. The geometry of a ray lies at the foundations of photogrammetry and therefore the geometrical aspects will be further discussed. This approximation is accepted when the wavelength of incident light is far shorter in comparison with the surface imperfections it encounters [Nayar et al. 1991].

The way light is scattered on the surface can be represented by mathematical models conceived physically (e.g. Cook-Torrance, Torrance-Sparrow, Kajiya, He-Torrance-Sillion-Greenberg), or empirically with the help of real samples (e.g. Strauss, Ward). A thorough review of available BRDF models was published in [Montes and Ureñ 2012]. The main distinction between the two models is in how accurate the surface structure is simulated, the physical being very precise, while the empirical only approximate, stressing the visual impressions, fast computation time, and universalism. Physical models

(i) regard materials at a molecular scale, (ii) comply with energy conservation law, and (iii) comply with the reciprocity principle (*Helmholtz stereopsis*). In other terms, incident light must be reflected or absorbed with no energy creation, whereas swapping the incident and exitance ray does not change the scattering distribution function. The equations that explain in amplitude what portion of the incident light is reflected and refracted are the so-called *Fresnel's formulas*.

Since in reality, even for the same material type, surface structures vastly differ, even for the same material type, recovering highly accurate models is seldom practised. Instead, simplified physical models are proposed. They reduce the problem by modelling the surface structure at a larger scale, while cautiously documenting the adopted assumptions.

Empirical models, on the other hand, respect the geometric principles of reflection and transmission, but surface appearance is steered with a set of parameters focused on delivering an optically pleasing result yet of little physical significance [Glassner 1995; Szeliski 2010; Weyrich et al. 2008].

BxDF Nicodemus et al. [1977] expressed the relationship between materials and light through a quantity that relates the incident flux $d\phi$ arriving from the differential solid angle $d\vec{\omega_i}$ at point $P(x_i, y_i)$, to the reflected radiance dL_r into the direction $\vec{\omega_r}$ (Figure 2.1). They called this generic function the *bidirectional scattering distribution function* (BSDF), underlining with the *bi-* prefix the dependence on incoming and outgoing directions; see S in Equation (2.1). The incident flux $d\phi$ is often represented by irradiance dE on a surface area dA as given below.

$$dL_r(\theta_i, \phi_i, \theta_r, \phi_r, x_r, y_r) = S(\theta_i, \phi_i, x_i, y_i, \theta_r, \phi_r, x_r, y_r) \cdot d\phi_i$$
$$= S \cdot dE_i \cdot dA_i \tag{2.1}$$

where $dE_i = L_i \cos\Theta_i \, d\omega_i$ and L_i is the incidence radiance; see Figure 2.1.

BSDF can be generalized to a sum of *bidirectional reflectance distribution function* (BRDF) and *bidirectional transmittance distribution function* (BTDF). The BRDF component covers all cases where the reflected light remains on the same side of the surface, as opposed to the BTDF component where the ray traverses to the counter hemisphere; see Figure 2.1. Adhering to planarity of the surface elements, uniform irradiation, uniformity and isotropy of the reflecting surface, gives way to a formulation of the scattering function which is independent of the point of observation. BSDF then becomes the BRDF — the proportionality function f_r as in Equation (2.2) — and is characteristic of reflecting surfaces. When the incident flux is totally reflected towards the specular direction hence $\theta_i = \theta_r$ and $\phi_i = \phi_r \pm \pi$, BSDF reduces to Equation (2.3); see the derivation in Nicodemus et al. [1977].

$$f_r(\theta_i, \phi_i, \theta_r, \phi_r) = \frac{dL_r(\theta_i, \phi_i, \theta_r, \phi_r, E_i)}{dE_i(\theta_i, \phi_i)} \tag{2.2}$$

$$f_r(\theta_i, \phi_i, \theta_r, \phi_r) = 2\delta \cdot (\sin^2\theta_i - \sin^2\theta_r) \cdot \delta \cdot (\phi_i - (\phi_r \pm \pi)) \tag{2.3}$$
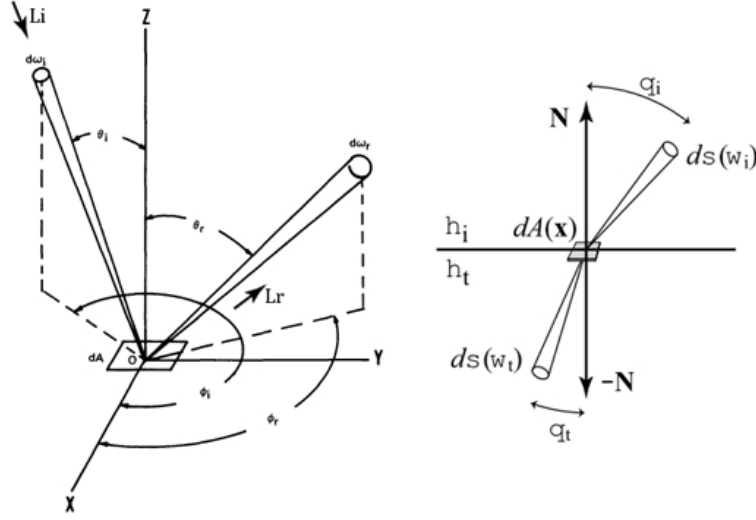
Figure 2.1: Geometry of the incident, reflected and transmitted beam [Glassner 1995].

where δ is the *Dirac* delta-function defined by

$$\delta(u) = 0 \quad u \neq 0,$$
$$\int \delta(u)du = 1$$
$$\int f(u)\delta(u-a)du = f(a);$$

(2.4)

If translucent objects are in question and transmission takes place, the BTDF must be evaluated. Because of the difference in density between two interfacing media, the linearity of the ray path when trespassing the object volume is violated. Deriving from Snell's law, it can be shown that the incident and exitant radiance are related by a ratio of squares refractive indices. The final BTDF written in terms of angles (θ, ϕ), as presented by Veach [1997], is given below. η_i and η_t are refractive indices of two different media, the δ function is an ordinary distribution function.

Modelling the appearance of an object, the BxDF function must be estimated for every point on the surface. For BRDF it then becomes the SVBRD, that is *Spatially-Varying BRDF*.

$$f_t(\theta_i, \phi_i, \theta_t, \phi_t) = \frac{\eta_t{}^2}{\eta_i{}^2} \cdot 2\delta \cdot (\sin^2\theta_i - \frac{\eta_t{}^2}{\eta_i{}^2}\sin^2\theta_t) \cdot \delta \cdot (\phi_i - (\phi_t \pm \pi))$$

(2.5)

Reflection is the process whereby light of a specific wavelength incident on a material is at least partly propagated outward by the material without change in wavelength [Glassner 1995]. The elementary models recognize few forms of reflection (see Fig-

9

ure 2.2), these being (i) specular, (ii) diffuse, (iii) glossy, (iv) mixed, and (v) retro-reflective. Specular reflection occurs when no surface scattering takes place, i.e. the case of smooth materials, with the total incident light reflected towards a single exitant direction. The outgoing and ingoing rays form the same angle with respect to the surface normal, plus they occupy a plane thus are coplanar, which altogether formulates the *law of reflection*; see Figure 2.2a.

Diffuse (the *Lambert* model) surfaces in an idealised example reflect light equally in all directions, hence are isotropic and have a constant BRDF. Two mechanisms that give rise to diffuse reflection are multiple scattering on the surface and internal scattering. Owing to this, it is often called the *body reflection*, while specularity is called the *surface reflection* [Nayar et al. 1991]. In reality, such materials do not exist. The materials we are familiar with are by and large a union of diffuse and specular reflection (as the *Phong*, *Torrance-Sparrow* models affirm). They are also the kind of materials one can measure remotely with least effort because the object's appearance is indifferent to the viewing position. Forms in c and d of Figure 2.2 are standarized instances that describe them.

Last, but not least, the retro-reflectives return the complete signal back in the direction of the emitter. Retro-reflecting sheets are especially popular in industrial close range photogrammetry applications where high precision and automation are put in the first place. The strong return in the direction of the camera (i) allows for the separation of relevant from irrelevant information in the object space assuring higher reliability, and (ii) diminishing the light fall-off at the targets' edges granting a better centroiding accuracy on that point.

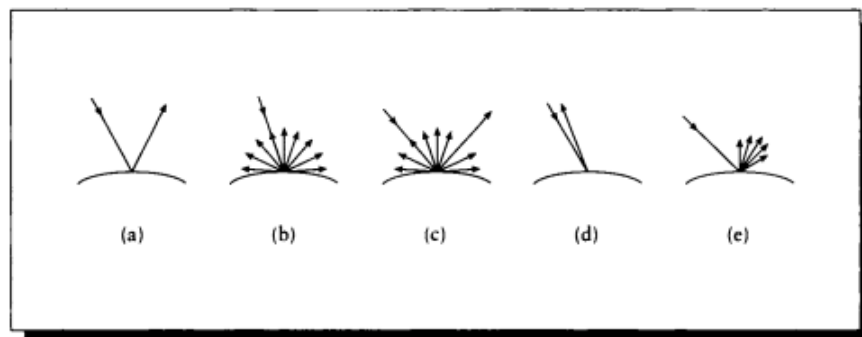**Figure 2.2:** Different reflection forms. (a) Specular, (b) diffuse, (c) elim, (d) retrore-flection, (e) glossy. Adapted from [Glassner 1995].

Refraction is the process whereby light of a specific wavelength incident on the interface (or boundary) between two materials passes (or refracts) through the interface and into the other material without change in wavelength [Glassner 1995]. Refraction

models are analogous to the reflection models of the previous section with the exception that they occur on the other side of the surface and reproduce the characteristic bending effect; see Figure 2.3. This bending effect is a consequence of different media densities and the fact that the light's phase velocity varies from medium to medium. The amount of bending is captured by the indices of refraction of two interfacing materials. The relation was coined by Snell, and is now known as the *Snell's law* — Equation (2.6). It is worth noting definition of individual refractive indices which is the ratio of the speed of a particular wavelength in vacuum to its speed in the medium on hand; see Equation (2.7).

$$\eta_i \cdot \sin\theta_i = \eta_t \cdot \sin\theta_t \tag{2.6}$$

$$\eta_\lambda = \frac{c}{v_\lambda} \tag{2.7}$$
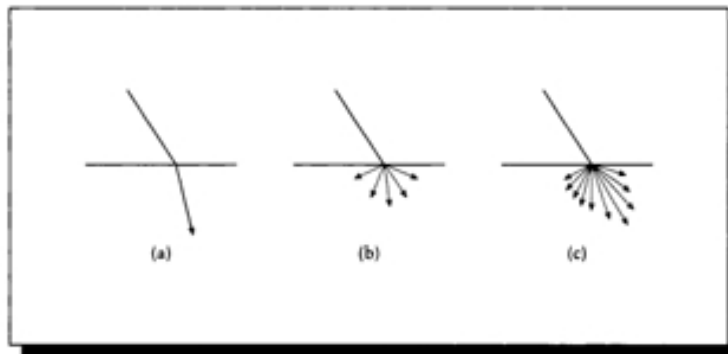
Figure 2.3: Different refraction forms. (a) Specular, (b) diffuse, (c) mixed. Adapted from [Glassner 1995].

2.2 Camera geometry

The mapping between 3D world space and 2D image space with a perspective camera is realised by the *central projection*; see Figure 2.4. The 3D point P and the center of projection O, constitute a spatial ray that intersects the 2D image plane in P'. The mathematical model relating image observations with the physical objects derives from *spatial similarity transformation*. It transforms objects between global (object space) and local (image space) three-dimensional Cartesian coordinate systems by translation, rotation, and scaling

$$\mathbf{x} = \lambda R^T(\mathbf{X} - \mathbf{X}_0) \tag{2.8}$$

Figure 2.4: Central projection relates the physical object with its points in the image plane [Kraus 1997].

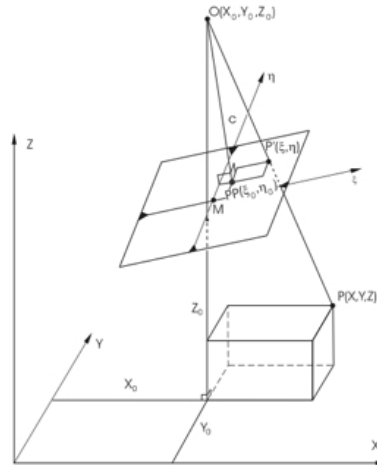which in essence brings the origins of the camera[1] and object coordinate systems to coincide, and rotates the two systems to be mutually parallel. If one decides to work in projective space, a linear transformation takes place

$$
\begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = \begin{bmatrix} \lambda R & -\mathbf{X}_0 \\ \mathbf{0}^t & 1 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix}.
\tag{2.9}
$$

Unlike in computer vision, where practical problems are mainly solved in projective space using homogeneous coordinates, in photogrammetry, Euclidean geometry and Cartesian coordinates are preferred[2].

Irrespective of parametrization, (i) whether in Euclidean or projective for 3D points, (ii) using Euler angles or quarternions for rotations, the problem of reconstruction is non-linear in its parameters. Least squares adjustment — the fundamental tool for parameter estimation in photogrammetry and across many other disciplines; see Chapter 3 — expects that the observations are functions of linear parameters. Consequently, it imposes that the imaging equations are linearized, hence their approximate values must be known a priori and the solution reached iteratively.

The collinearity equation and the projection matrix that are discussed in the following paragraphs offer one way to recover interior/exterior orientations of the cameras or perform intersections i.e. shape reconstruction. An alternative mathematical concept is the coplanarity condition, also known as the fundamental or essential matrix. Subsection 2.3.2 introduces this alternative concept in detail.

---

[1]It is assumed that the origin of the camera coordinate system coincides with the center of perspective.
[2]In computer vision community also known as projective and euclidean reconstructions.

Euclidean geometry The familiar way to put the spatial similarity relation is in the form of *collinearity equations*, obtained by grouping observations on the left hand-side, free and fixed parameters to the right, and removing the scale factor by dividing the x and y rows with the z; see Equation (2.10). The principal point and principal distance are denoted as — x_0, y_0, c, and define parameters of the camera coordinate system (camera model or inner orientation). Additional distortion parameters to correct for the systematic errors introduced by the lens, which are inherently non-linear, are denoted as $\Delta x, \Delta y$. The X_0, Y_0, Z_0 coordinates of the camera projection center and r_{ij} elements of the rotation matrix[3] R determine the position and orientation of the camera coordinate system (exterior orientation) expressed in the coordinate system of the object. The inverse of the collinearity equation allows for the computation of planimetric coordinates of a point lying on a plane of constant Z; see Equation (2.11).

$$
\begin{aligned}
x - x_0 + \Delta x &= -c\frac{(X - X_0) \cdot r_{11} + (Y - Y_0) \cdot r_{21} + (Z - Z_0) \cdot r_{31}}{(X - X_0) \cdot r_{13} + (Y - Y_0) \cdot r_{23} + (Z - Z_0) \cdot r_{33}} \\
y - y_0 + \Delta y &= -c\frac{(X - X_0) \cdot r_{12} + (Y - Y_0) \cdot r_{22} + (Z - Z_0) \cdot r_{32}}{(X - X_0) \cdot r_{13} + (Y - Y_0) \cdot r_{23} + (Z - Z_0) \cdot r_{33}}
\end{aligned}
\tag{2.10}
$$

$$
\begin{aligned}
X - X_0 &= (Z - Z_0)\frac{(x - x_0 + \Delta x) \cdot r_{11} + (y - y_0 + \Delta y) \cdot r_{12} + (-c) \cdot r_{13}}{(x - x_0 + \Delta x) \cdot r_{31} + (y - y_0 + \Delta y) \cdot r_{32} + (-c) \cdot r_{33}} \\
Y - Y_0 &= (Z - Z_0)\frac{(x - x_0 + \Delta x) \cdot r_{21} + (y - y_0 + \Delta y) \cdot r_{22} + (-c) \cdot r_{23}}{(x - x_0 + \Delta x) \cdot r_{31} + (y - y_0 + \Delta y) \cdot r_{32} + (-c) \cdot r_{33}}
\end{aligned}
\tag{2.11}
$$

The collinearity condition can serve a range of purposes, when employed as the functional model of the imaging process. Depending on which part of the equation is used as observed/measured, constant and unknown, resolved are the tasks of

- camera calibration ($x_0, y_0, \Delta x, \Delta y$ as unknowns or observed unknowns),

- image orientation ($R, X_0 Y_0 Z_0$ unknown or observed unknown),

- point intersection (XYZ unknown),

- point back-projection (x, y unknown).

Difficulties in formulating mapping using Euclidean geometry include (i) the translation being a non-linear operation precluding the simplistic expression of the imaging equation as a matrix product, and (ii) no direct methods existing to solve the general task of image orientation when no camera model is given.

---

[3] There exist numerous parametrizations of the three elementary rotations; parametrization is always local and chosen to avoid singularities; in close-range photogrammetry the α around Z, ζ around Y, κ around Z prevails.

Projective geometry Reconstruction in the computer vision community is equivalently performed using projective geometry, and results in a reconstruction up to projective transformation[4]. Euclidean reconstruction is also practised, however, it is always preceded by the projective restitution and finally upgraded to the Euclidean world, provided that the camera model is known [Hartley and Zisserman 2003]. In the common language, the *pose* or *motion* estimation stand for the image orientation, whereas the *scene* or *structure* computation refer to the intersected 3D points.

The field of computer vision values efficient and automated processing that can handle large amounts of data. Because the reconstruction problem is highly non-linear in its parameters — orientations, object coordinates, possibly distortion parameters, and other constraints — the type of parametrization that will avoid singularities and act reliably plays a crucial role. Rotations, points, lines and planes do not have a global parametrization, therefore to assure a numerically stable behaviour, reconstruction is performed (i) incrementally (non-linear least squares; true also for Euclidean geometry; see Chapter 3), (ii) in quaternions, and (iii) embedded inside the projective framework. This geometry is straight line preserving and describes shapes and angles in the same way that Euclidean space does, except that points at infinity become full-fledged, from now on called the *ideal points*. Triggs et al. [2000] argue that for this property, the problem of reconstruction is intrinsically projective. As an example, the authors say that distant 3D points require large displacements in order to act on the image, whereas with homogeneous coordinates, points near infinity have no distinction and act as naturally as close-range points.

This feature originates from the fact that homogeneous entities (point, line, plane) that describe the geometry are invariant to multiplication by a scalar $w \neq 0$, so $w\mathbf{r} = \mathbf{r}$. A point in 2D Euclidean space (x, y), is represented in projective space by the Euclidean and homogeneous part (in Equation (2.12)). By moving from one space to another, a division by the scalar is performed, but when $w = 0$ the Euclidean part becomes infinite while the projective representation takes the form of $[u, v, 0]$, which is the point at inifity [McGlone 2004].

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_0 \\ x_h \end{bmatrix} = \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} wx \\ wy \\ w \end{bmatrix} \tag{2.12}$$

Any Euclidean transformation can be written in the form indicated in Equation (2.13), where $\mathbf{w}$, $\mathbf{v}$ are vectors, and the M matrix is 2 x 2 (planar case), of a given type (representing rotation or reflection). This formulation is somewhat awkward to handle because the translation is not a linear operation and is additive. Implementing inverses or compositions of such a transformation in a computer environment becomes an effort. Yet, if a Euclidean point is given in homogeneous coordinates $(x, y, 1)$, translation takes the more compact form of Equation (2.14) and is applied, similarly to rotation and

---

[4]Projective ambiguity implies that points $\mathbf{X}$ in the scene and projection matrix P can be multiplied by another 4x4 matrix H without changing their reprojections to image points i.e. $P\mathbf{x} = (PH^{-1})(H\mathbf{X})$.

14

reflection, by multiplication [Richter-Gebert 2011] .

$$\mathbf{p} \to M(\mathbf{p} - \mathbf{v}) + \mathbf{w} \tag{2.13}$$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{2.14}$$

This representation of coordinates brings about the compressed formulation of the imaging equation, i.e. the rotation matrix and translation vector can be multiplied and represented together inside the so-called *projection matrix*; see P in Equation (2.15). It is a 3 x 4 matrix containing 11 parameters[5] (and thus 11 *degrees of freedom*), 5 of which refer to the calibration matrix (interior orientation) and the other 6 describe the camera pose (exterior orientation). The individual parameters can be obtained through decomposition (e.g. eigenvalue computation and singular value decomposition SVD) of the P matrix into a product of K, R and O that represent respectively the calibration matrix and the rototranslation between the camera coordinate frame and the world frame [Hartley and Zisserman 2003]. The 5 elements of the calibration matrix are (i) x_0, y_0 as the principal point, (ii) c_x, c_y as the focal length in x and y directions, and (iii) s as the skew. Equation (2.17) rewrites the projection matrix in terms of elements p_{ij} giving rise to the known direct linear transformation (DLT) [Abdel-Aziz and Karara 1971]. The discussion of DLT continues within Section 2.3.1.

$$\begin{aligned} \mathbf{x} &= \begin{bmatrix} c_x & s & x_0 \\ & c_y & y_0 \\ & & 1 \end{bmatrix} \mathbf{X}_{cam} \\ &= \begin{bmatrix} c_x & s & x_0 \\ & c_y & y_0 \\ & & 1 \end{bmatrix} \begin{bmatrix} R & -RO \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \\ &= KR[I| - O]\mathbf{X} = P\mathbf{X}, \end{aligned} \tag{2.15}$$

A note on the camera model The camera, from the perspective of interior orientation, can be distinguished as *general*, *straight-line-preserving*, and *calibrated*. Many photogrammetric cameras are denoted as calibrated due to their orientation known up to high accuracy. The straight-line-preserving (aka projective, affine or uncalibrated) cameras are those that preserve the straight lines in the imaging process and are therefore free of lens distortion, yet other than this, their geometry is unknown. Interior orientation of the general cameras are completely unknown. If a straight-line-preserving camera model is assumed, the resulting reconstruction is a projective reconstruction, and lastly, a calibrated camera results in Euclidean reconstruction. For further insights see the paragraph below and Subsection 2.3.2.

---

[5]A 3 x 4 transformation matrix applied to a projective entity has only 11 DoF because the entity is defined up to a non-zero scalar.

A note on 3D reconstruction ambiguity. The outcome of 3D reconstruction can be (i) projective (PR), (ii) affine (AR), or (iii) Euclidean (metric, ER). The type of reconstruction derives from the assumed camera model and implies certain ambiguities. They differ respectively by projective, affine and similarity transformation with respect to the true reconstruction. For the PR, the invariant properties are intersections and surface tangency. For the AR invariant parallelism of planes is additionally assumed, whereas the ER preserves both the parallelism and the true angles. Further consequence of the particular invariance is that the result is ambiguous up to a given class of transformations. For instance, in the PR, any multiplication of the result by a 4×4 matrix will leave the result untouched.

Each type of reconstruction can be upgraded to the metric result. Note that the ER $P_E, P'_E, \mathbf{X}_{Ei}$ and PR $P, P', \mathbf{X}_i$ (as well as PR and AR or AR and ER) are related by a non-singular matrix H such that

$$
\begin{aligned}
P_E &= PH^{-1} \\
P'_E &= P'H^{-1} \\
\mathbf{X}_{Ei} &= H\mathbf{X}_i.
\end{aligned}
\tag{2.16}
$$

In order to compute the H matrix and transform the result to the metric one, one may exploit the ground truth information or move from projective to affine to Euclidean reconstruction by imposing a number of constraints. The general reconstruction workflow is as follows [Hartley and Zisserman 2003]

- Compute projective reconstruction $P, P', \mathbf{X}_i$

  - Compute the fundamental matrix F from point correspondences (see Subsection 2.3.2)

  - Retrieve the cameras P, P' from the fundamental matrix

  - Perform 3D intersection

- Upgrade the projective to metric reconstruction

  - The direct method, with the use of ground control points

  - The stratified method

    a Affine reconstruction. The camera with perspective center at infinity. Parallel lines will be projected to the camera as parallel lines. Find the transforming homography H by identifying the plane at infinity by e.g. introducing constraints on parallel lines.

    b Euclidean reconstruction. Introduce constraints such as scene orthogonality, or constraints on camera interior orientation to solve for the H matrix.

16

2.3   Diffuse surfaces

Inferring shape from images acquired with passive techniques is possible thanks to visual cues such as (i) stereo disparity, (ii) shading, (iii) focusing, and (iv) texture [Szeliski 2010]. Besides the diffuse surface, the prerequisites to carry out shape measurements are correspondingly (i) homologous features for the stereo principle, and/or (ii) one or more point light sources for the shading, and/or (iii) the minimum of two blurred images focused at different depths to derive the depth from defocusing, and/or (iv) repetitive patterns.

Continuing the discourse of Section 2.2 on camera geometry, the coming sections discuss other mathematical instruments that tie image information with physical reality. Neither *shape from texture*, nor *shape from focus* are given insight in the following but more on the subject can be found in Wolff et al. [1992]. The text is structured into single and multi-view approaches. The camera model is assumed perspective except in *shape from shading,* and *photometric stereo* where it is generalized to the orthographic projection.

2.3.1   Single view

Shape reconstruction from a a single image in the most generic case is impossible. When the three dimensional world is projected to a two dimensional image, one dimension is lost. The 2D map captured in the image cannot revive the third dimension unless additional constraints are imposed on the object under reconstruction. This loss of information is inherent to all passive imaging techniques, i.e. those recording and exploiting merely intensity data for the purpose of metrology. Practically speaking, the back-projection of a point $P(X, Y, Z)$ to a point $p(x, y)$ in the image is doable provided camera orientations are given (Equation (2.10)). For the reverse operation, as the Equation (2.11) depicts, unless one of the three coordinates is fixed, there exists a line of solutions.

Camera orientation parameters, as incidentally remarked upon numerous times before, is a prerequisite in any reconstruction task. For this reason, subsequent paragraphs will comment on plausible ways to retrieve these parameters with a single camera, interleaving it with notes on constraints to be applied to make the measurement possible. We start with spatial resection and the perspective-n-point pose algorithm. Next, we proceed to a more general case of simultaneous orientation and calibration[6] of a camera. Lastly, reported are (i) the specialized case when 3D points lie on a plane, (ii) the vanishing points, lines and the information they carry, and (iii) two methods that rather than exploit geometry, use mathematical models of surface reflection to build its 3D representation. Methods from photogrammetry and computer vision are be recalled.

Spatial resection, perspective-n-point pose   Estimating the position and orientation of a single image is a deep-rooted problem in the field of photogrammetry and com-

---

[6]Calibration in this sense shall be understood as inner orientation

puter vision. The nomenclature used within the fields is different, *spatial resection* and *perspective-n-point pose (PnP)* respectively, but the problem is defined uniquely: given a set of correspondences between the image and 3D object, all in their proprietary coordinate systems, find the position of the center of perspective and orientation of the camera expressed in the object frame. Spatial resection was the first orientation procedure[7] to be applied and Grunert [1841] was the first to propose a closed-form solution with the minimum number of points, i.e. three. Since then, numerous approaches to finding single image orientation have been conceived, either for perspective or projective camera models. Spatial resection is important and still widely used today. In combination with intersection, it allows for the automatic building of the initial global exterior orientation of a block of $2+$ images. Every new image to be added to an existing block is resected, exploiting the 3D points intersected from the already-oriented set of images.

There exist 3-point and 4-point algebraic derivations of the problem and they always end up with a with a fourth-degree polynomial. Among photogrammetrists the accepted approach is by Killian [1955] because of its simple formulae and the unique solution that it delivers, even when all points lie on a plane or the critical cylinder [Huang 2001; Kraus 1997]. In computer vision, Fischler and Bolles [1981] developed the most popular non-linear P3P algorithm.

The above methods although closed-form are non-linear, and because they do not take the full advantage of the redundancy when $3+$ points are at hand, they are susceptible to noise [Quan and Lan 1999]. Iterative approaches are possible though seldom applied for the trouble of proper initialisation. The computer vision community found ways to present the problem in a linear way for any number of corresponding points (hence the perspective-n-point prefix). No further details to the plethora of computer vision techniques is presented here, however, the reader is referred to the following literature: Fischler and Bolles [1981]; Horaud et al. [1997]; Quan and Lan [1999]. Implementation of some PnP algorithm is available within the *Open Source Computer Vision Library (OpenCV)*.

Direct Linear Transformation Previous paragraph presented how to derive the three angles and position of a camera in 3D having the calibration parameters provided. There are situations when this piece of information is either unavailable or of less importance than the cost of the computation that it imposes.

The invention of the *Direct Linear Transformation* (DLT; Equation (2.17)) is attributed to Abdel-Aziz and Karara [1971]; Sutherland [1974]. The name DLT is interchangeably used with *projective transformation* or simply *homography*. The DLT directly relates points in object space to points in the images. It may work as resection or intersection, no different than the collinearity equation; see Equation (2.10). But as the name suggests, the method was meant to be purely linear so the non-linear camera distortion

---

[7]Spatial resection and PnP work with calibrated cameras. The direct linear transform that follows in the next paragraph serve the uncalibrated case.

parameters must be ignored. Secondly, the coefficients p_{ij} in

$$x - x_0 = \frac{p_{11} \cdot X + p_{12} \cdot Y + p_{13} \cdot Z + p_{14}}{p_{31} \cdot X + p_{32} \cdot Y + p_{33} \cdot Z + p_{34}}$$
$$y - y_0 = \frac{p_{21} \cdot X + p_{22} \cdot Y + p_{23} \cdot Z + p_{24}}{p_{31} \cdot X + p_{32} \cdot Y + p_{33} \cdot Z + p_{34}}$$
(2.17)

that are the elements of the homogeneous projection matrix P from Equation (2.15), must be treated as parameters, rather than functions of parameters that actually model the imaging process. This kind of formulation is particularly attractive as it (i) opens up a way to estimate the camera pose without knowledge of the camera behind the imaging process, and (ii) is linear in concept, and therefore offers efficient calculations and wider perspective for automation procedures. The myriad advantages are naturally counterbalanced by some disadvantages. In the first place, the beauty of the linear formulation comes at the expense of the number of 3D points preconditioned to solve for the camera pose (3 in Euclidean and 6 in projective geometry). DLT is parametrized with 11 elements, whereas in the conventional form (Cartesian coordinates) only 9 parameters are needed to describe the central projection. To interpret the two superfluous parameters in Euclidean space, one can think of a shear factor between x and y image axes, or else an affine transformation applied to the image coordinates[8]. To bear in mind is that the additional parameters are not linearly independent which may lead to an incorrect set of parameter values still nicely fitting the observations to the model.

Nonetheless, the linear formulation comes in useful when reconstruction from paintings or postcards is to be performed. And above all, as the algorithm thinks "images" rather than "cameras" by assining a set of unique DLT parameters to each image, it is capable of reconstructing scenes from collections of images of unknown origins and specifications.

As soon as one wants to include distortion into the DLT model, or input known interior orientation values, it adds non-linearity to a linear formulation by definition and ought to be solved iteratively by some non-linear least squares method.

2D homography When a 2D-3D mapping can be generalized to a 2D-2D mapping, one speaks of *planar homography*, sometimes also referred to as *2D projective transformation* or simply DLT. Indeed, the only distinction from the DLT from the previous section is in the regarded dimension. The 2D homography relates corresponding points situated on planes. It is defined by a $3x3$ matrix (see Equation (2.18)) over homogeneous coordinates, it is linear in the parameters of the unknown H matrix, and a minimum of 4 points are required for the solution (with e.g. SVD for homogeneous, or Gauss elimination for non-homogeneous result). When more points are available, the consistent solution must

---

[8]See Dermanis [1994]; Förstner [1999] or Kraus [1997] for relating the two parameter sets.

minimize a cost function (e.g. linear least squares methods). The matrix has a rank 8, therefore the results are up to a scale factor.

$$\mathbf{x}' = H\mathbf{x} \tag{2.18}$$

Cases in which the planar homography is useful include, when (i) relating two camera views with each other, (ii) relating a plane in the scene with the camera view, (iii) estimating camera rotation from homography, or (iv) performing mosaic stitching.

The epipolar geometry (see Subsection 2.3.2) is one way to constrain the relation between a point in an image and a line being its projection in another image. Next to the epipolar lines, the planar homography can map points in one view, to points in another view by projecting to that view a 3D point, which arises from the intersection of the first view ray with a plane. To illustrate the dependencies, suppose there are two cameras with their camera matrices $P = [I|\mathbf{0}]$ and $P' = [R|\mathbf{t}]$, such that $\mathbf{x} = P\mathbf{X}$ and $\mathbf{x}' = P'\mathbf{X}$. There is also a plane in front of them, such that it does not contain the camera centers and is defined as $\pi\mathbf{X} = 0$ where $\pi = (\mathbf{v}^T, 1)^T$. Following Equation (2.18) and the above, the plane induced homography equals $H = R - t\mathbf{v}^T$, and points of the first camera project to the second camera as $\mathbf{x}' = (R - t\mathbf{v}^T)\mathbf{x}$.

Moreover, the need to map the image plane with the scene plane crops up when a planar object, e.g. a building edifice, is taken from an oblique angle, yet an orthogonal view is solicited. To produce a view free of the perspective distortion, such that will simulate an image taken parallel to the object, one must retrieve the 2D homography. As noted before, merely 4 points suffice for the operation and no camera parameters must be known.

Another two interesting examples where 2D projective dependencies that alleviate the measurements are that of a camera undergoing a pure rotation, and a camera whose focal length changes yet the perspective center remains the same. In the latter scenario, a 3D point is projected to the camera before and after zooming as $\mathbf{x} = K[I|\mathbf{0}]\mathbf{X}$ and $\mathbf{x}' = K'[I|\mathbf{0}]\mathbf{X} = K'K^{-1}(K[I|\mathbf{0}])\mathbf{X} = K'K^{-1}\mathbf{x}$. That is, the homography in Equation (2.18) is replaced with $H = K'K^{-1}$. In a similar manner, when pure rotation takes place the $\mathbf{x}$ remains the same while $\mathbf{x}' = K[R|\mathbf{0}]\mathbf{X} = KRK^{-1}K([I|\mathbf{0}])\mathbf{X} = KRK^{-1}\mathbf{x}$, and the homography is of the form $H = KRK^{-1}$. The pure rotational case is commonly exploited in creating panoramic views where a set of images is acquired from a camera centered above a single point.

For more information, the reader is directed to Hartley and Zisserman [2003]; Mc-Glone [2004]; Szeliski [2010], which served as the reference material for this paragraph.

Vanishing points Vanishing points occur in images as intersections of image lines coming from parallel 3D lines. For instance, parallel contours of a building are rendered convergent in the image space and meet in the vanishing point $\mathbf{v}$ (see Figure 2.5). Thus, the vanishing point is common for lines of same directions, and at the same time indifferent to their position.

Let's say a line in 3D space passes through a point $\mathbf{A}$ and follows the direction $\mathbf{D} = (\mathbf{d}^T, 0)^T$. It is then expressed in parametric forms as $\mathbf{X}(\lambda) = \mathbf{A} + \lambda\mathbf{D}$, with the position of
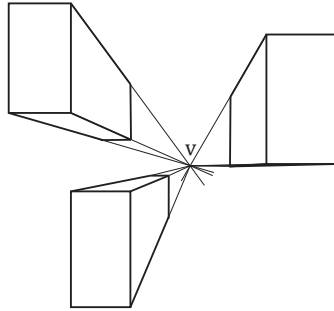
Figure 2.5: Lines parallel in 3D space meet at the vanishing point **v** in 2D image space.

the point changing according to the λ parameter. The vanishing point can be obtained from the λ value as a limiting point (see Figure 2.6, Equation (2.19)), that given the camera as $P = K[I|\mathbf{0}]$ projects to Eqaution 2.20. The projected points depend on K being the camera calibration matrix and **d** being the 3D line direction, which again seconds that the projection remains independent of the position.
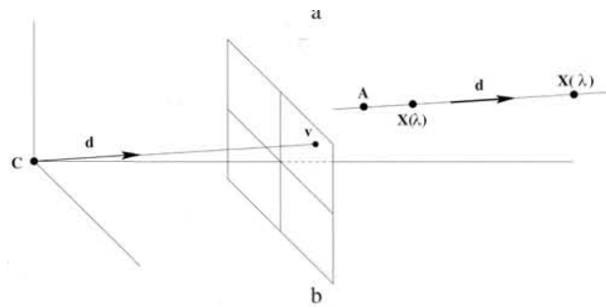

Figure 2.6: Projection of a point at infinity $\mathbf{X}(\lambda)$ through the camera center **C** to a vanishing point in the 2D image plane. Source: Hartley and Zisserman [2003].

$$\mathbf{v} = \lim_{\lambda \to \infty} \mathbf{x}(\lambda) \tag{2.19}$$

$$\mathbf{v} = \lim_{\lambda \to \infty} P\mathbf{x}(\lambda) = \lim_{\lambda \to \infty} (\mathbf{a} + \lambda K\mathbf{d}) = K\mathbf{d} \tag{2.20}$$

This somewhat abstract formulation turns out useful in practical applications. One may assume there is a line in 3D space that is observed from different viewpoints by two cameras. One has then two vanishing points $\mathbf{v}$, $\mathbf{v}'$ in two cameras, hence two line directions $\mathbf{d}$, $\mathbf{d}'$, correspondingly. Taking into account the relation of Equation (2.20),

the two camera viewpoints are linked by the two directions and the rotation of one with regard to another. The rotation is thence possible to recover. Vanishing points may also serve to restore the camera focal length or even polyhedral objects embedded in the 3D scene. For more information the reader is directed to Hartley and Zisserman [2003]; McGlone [2004]; Szeliski [2010], which served as the reference material for this paragraph.

Shape-from-shading and photometric stereo. Thus far the imaging process has been regarded in the geometric perspective, but radiometry, that is the gradual shading of the image, likewise carries information on the object shape. For a surface, some illumination source and a fixed imaging geometry, one can compute a reflectance value for every possible orientation of that surface normal. This information is captured within the so-called reflectance map R. It further means that brightness I observed in an image (i.e. image irradiance) is also a function of surface gradient (i.e. orientation), and under distant light, distant observer is equal to the computed reflectance

$$I(x,y) = R(p,q,x,y) \approx R(p,q). \tag{2.21}$$

The assumption of the distant light and the observer allows for the generalization of the underlying reflection model to be independent of the viewing direction (orthographic projection). For every image point there will exist only one reflectance map. Consequently, the shape-from-shading techniques will provide erroneous results on shiny surfaces characterized by specular reflections.

The relationship between surface brightness and its gradient recorded in a reflectance map was introduced by Horn [1977]. The author described the shape of the surface in the camera-centered coordinate system i.e. the z axis runs along the camera optical axis, the xy-plane is fixed at a reference depth and coincides with the coordinate system of the image. Surface gradients are then expressed with pq coordinates defined as $p = \partial z/\partial x$, $q = \partial z/\partial y$, so it is the rate of change of depth that constitutes the *gradient space*. Distance from the origin within the pq-plane reflects the slope, whereas the direction is the steepest ascent [Ikeuchi and Horn 1981].

Ikeuchi and Horn [1981] proposed an evocative way to represent all possible surface orientations on a unit *Gaussian sphere*; see Figure 2.7. Imagine the viewer placed above the northern pole and the z extending from the center of the sphere towards the north. To map orientation of a surface patch onto that sphere, it is "moved" to its center, with no change of attitude in space. Then the spot on the sphere that is faced by the normal of that patch is used to identify that particular orientation. The authors generalize the projection to an orthographic one which allows them to state that objects patches sharing the same orientation are represented by a single point on the *Gaussian sphere*, irrespective of their placement on the object. The final reflectance map is represented by 2D stereographic projection of the 3D sphere; see Figure 2.7 (left).

Establishing the reflectance map is relatively easy. One needs to model the scene with a fixed illumination source, imaging geometry and surface properties. Assigning brightness values to specific surface orientations (i.e. spots on the Gaussian sphere)
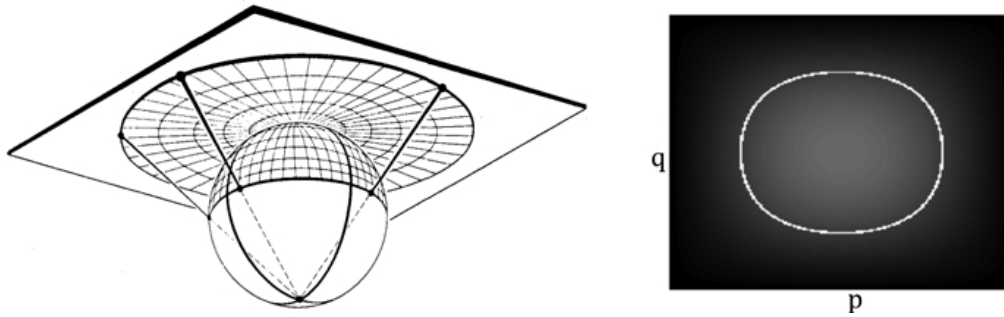
22

Figure 2.7: Left: Stereographic mapping of the *Gaussian sphere* [Ikeuchi and Horn 1981]. Right: An examplar reflection map of a lambertian surface; the white contour corresponds to a set of orientations consistent with one image brightness [*James P. O'Shea.*].

is straight-forward and can be carried out (i) experimentally using a sample object mounted on a goniometer, or (ii) theoretically using the *Bidirectional Reflectance Distribution Function BRDF*; see Section 2.1.

Finding the right surface orientation is by far more challenging, because it cannot be resolved with a single brightness measurement. The image data merely fixes the relationship between p and q, and narrows down the possible solutions to a set of surface orientations represented by the selected brightness; see the white contour in Figure 2.7 (right). The shape from shading formulated in Equation (2.21) is itself nonlinear when parameterized with the surface gradient pq. For a complete surface retrieval, additional constraints are required. A thorough survey of the shape from shading methods was performed by Zhang et al. [1999], according to whom, the techniques can be divided into those (i) minimizing a global energy function, (ii) propagating shape information from several seed points, (iii) assuming a surface type locally, and (iv) linearizing the reflectance map.

The same survey states that shape from shading conducted on real images delivers unsatisfactory results due to the generalization of perspective to orthographic projection and too simplistic modelling of the surface properties with the Lambertian model. It is suggested that the methods should employ more realistic surface models, accounting for specular/glossy reflections and combined with other cues such as stereo, range or multiple image data. An examplar shape from shading reconstruction with an assumption of local surface is shown in Figure 2.8.

The underconstrained problem of determining shape from shading can be overcome when the scene is imaged twice from the same view but with two different illumination sources. Two reflectance maps are produced and every image pixel (i.e. brightness or image irradiance) can be associated with two contours in the respective gradient spaces. Intersecting them may uniquely determine the orientation of the object point corresponding to the image pixel, and if not, a third image should resolve the ambiguity.
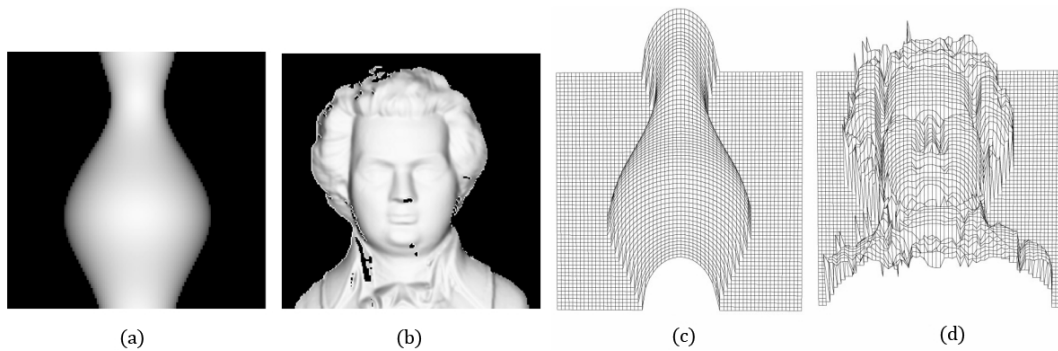
Figure 2.8: (a-b) Synthetic images of a vase and Mozart. (c-d) Shape from shading result solved with a local approach [Zhang et al. 1999].

The method was introduced by Woodham [1980] and is called the *photometric stereo*.

2.3.2 Multi-view

The single view shape reconstruction is generally ill-posed unless assumptions on the 3D scene are imposed. This is a great deficiency of the approaches discussed in the previous section. Employing more views permits unambiguous reconstructions while relaxing the extra assumptions. Subsequent paragraphs will comment on geometrical constraints that exist between two or more camera views.

2-view geometry. The geometry of an image pair is fully described by absolute orientation or with the projection matrices P with 6 or 11 parameters, depending upon whether the camera is calibrated or straight-line-preserving (see Equation (2.15)). The relative orientation of an image pair is described in terms of epipolar geometry telling that homologous image points $\mathbf{x}, \mathbf{x}'$, the object point $\mathbf{X}$ and the two camera centers are coplanar. There are 5 or 7 unknown parameters, whether the camera is calibrated or straight-line-preserving. The condition is encapsulated inside the coplanarity equations, fundamental or essential matrices. All three, up to slight variations, encode the same information.

Given the condition and a known image point $\mathbf{x}$, one may ask how the point $\mathbf{x}'$ in the other image is constrained. The two camera perspective centers form a baseline that together with the ray, defined by $\mathbf{x}$ constitute a plane π. The point $\mathbf{x}'$ in the other image is therefore constrained to lie within the line of intersection of the plane π and the plane of the second image (see Figure 2.9). The established line $\mathbf{l}'$ is called the *epipolar line* and is of great significance in all matching algorithms, as it constrains the search for corresponding points from 2D to 1D space.

The relationship between an image pair is fully expressed with the triple scalar product (see Equation (2.22)). Three vectors, $\mathbf{p}, \mathbf{p}'$ and the baseline $\mathbf{b}$ form a parallelepiped
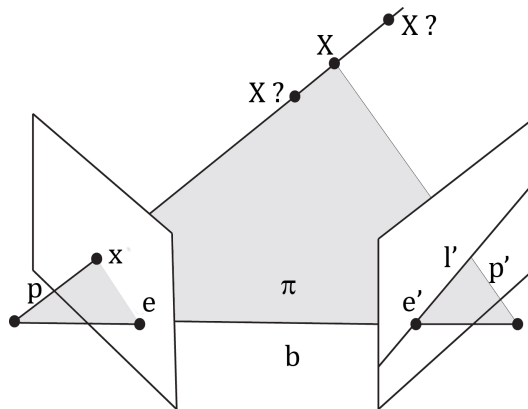
24

Figure 2.9: The coplanarity condition. Homologous points $\mathbf{x}$ and $\mathbf{x}'$, and the two perspective centers lie on a common epipolar plane. The epipolar line $\mathbf{l}'$ is the intersection of the epipolar and the image planes. Adapted from Hartley and Zisserman [2003].

whose volume is zero provided the vectors are coplanar. The $\mathbf{p}$, $\mathbf{p}'$ vectors are derived from image coordinates (x, y),(x', y'), the focal length c and rotation matrices R, R'. The rotations relate the cameras in the relative (local) coordinate system. Moreover, one may explicitly declare the cameras' interior orientations to be (x_{0i}, y_{0i}, c_i) (see Equation (2.23)) and reformulate the scalar triple product accordingly (see Equation (2.24)). Finally, the fundamental F and essential E matrix emerge. The matrices are equal if the camera interior orientation is known, as depicts Equation (2.24). In either case the coplanarity constraint is quadratic in the image coordinates and linear in the unknown elements of the matrices.

$$\mathbf{p}^T (\mathbf{b} \times \mathbf{p}') = \mathbf{p}^T B \mathbf{p}' = \mathbf{p}^T \begin{bmatrix} 0 & -b_z & b_y \\ b_z & 0 & -b_x \\ -b_y & b_x & 0 \end{bmatrix} \mathbf{p}' \tag{2.22}$$

$$\mathbf{p} = R \begin{bmatrix} x - x_0 \\ y - y_0 \\ 0 - c \end{bmatrix} = R \begin{bmatrix} 1 & 0 & -x_0 \\ 0 & 1 & -y_0 \\ 0 & 0 & -c \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = RC \begin{bmatrix} x & y & 1 \end{bmatrix}^T \tag{2.23}$$

$$\begin{aligned} \begin{bmatrix} x & y & 1 \end{bmatrix} C^T R^T B R' C' \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} &= \begin{bmatrix} x & y & 1 \end{bmatrix} F \begin{bmatrix} x & y & 1 \end{bmatrix}^T \\ &= \begin{bmatrix} x & y & 1 \end{bmatrix} C^T E C \begin{bmatrix} x & y & 1 \end{bmatrix}^T \end{aligned} \tag{2.24}$$

The conventional relative orientation with calibrated cameras has 5 degrees of freedom (DoF). In the absolute formulation, each camera has 6 parameters giving a total of 12 parameters for a pair. If one "subtracts" the similarity transformation of the local to

absolute system i.e. 7 parameters, 5 independent parameters of the relative orientation remain. The 5 DoF may be parametrized in a number of ways. Classically, the origin of the local coordinate system is fixed in the perspective center of one camera, and the rotation matrix of that camera set to identity. Coupling another camera to this defined coordinate system allows for the calculation of the three rotation angles and two components of the base vector (the third component serves to fix the scale of the established model).

When the calibrated camera is replaced with a straight-line-preserving camera, the fundamental matrix comes into play. Straight-line-preserving cameras are determined by the 11-parameter camera projection matrices (DLT, see Equation (2.15)). For a pair of images in the absolute coordinate system 22 parameters exist. "Subtracting" the 15-parameter 2D homography from the total number gives the 7 independent elements of the fundamental matrix.

The classical strategies for solving the fundamental matrix are the minimum case 7-point algorithm, or the normalized 8-point algorithm. As the fundamental matrix is singular (note that the B is skew-symmetric), computation with SVD is employed. For calibrated cameras the most popular algorithm is the one applied by Nistér [2003] and relies on 5 homologous points.

The fundamental matrix is the foundation for the projective reconstruction from 2 uncalibrated cameras. If appropriate constraints on the scene and/or camera are imposed, metric results are also plausible (see Paragraph 2.2).

For 3-view and 4-view geometry, similar arguments can be brought to establish the relationships between views. One then speaks of the *trifocal* and *quadrifocal tensor*. Its derivation is much more involved due to the arising internal constraints, it therefore exceeds the scope of this review. In fact, quadrifocal tensors are seldom used in practical applications. Besides this, the multiplicity of solutions for the fundamental and essential matrix, as well as critical configurations that hinder precise recovery are also outside the scope of this work. For more information, the reader is referred to Hartley and Zisserman [2003]; McGlone [2004]; Ressl [2003], which above all served as bibliography for this paragraph.

N-view geometry. There are essentially two common techniques for reconstructing from multiple views

- Sequential, non-optimal, that reconstructs the scene bit by a bit using the 2-view or 3-view geometry, and

- Bundle adjustment (BA), optimal, where the cameras and the structure are simultaneously recoverd from multiple views.

The BA is, without doubt, the dominant methodology. It assumes a model that associates the observations with the unknown parameters. The observations in the context of 3D reconstruction are the image coordinates, while the parameters are the cameras

and structure. This particular model, as has been pointed out in previous Sections, is encapsulated inside the (i) fundamental and essential matrix, or (ii) collinearity equation and camera projection matrix.

Finding the optimal solution is equivalent to obtaining the model parameters that will minimize some cost function. Typically the problem is formulated as a linear or non-linear least squares adjustment (see Chapter 3) and the cost function is quadratic in the image re-projection error. A cost function outlined in this way ascertains that the estimated model parameters will generate data that is as close as possible to the observed data.

The BA is a global reconstruction method, and is non-linear. The ultimate result is reached iteratively, meaning that an arbitrary starting point does not guarantee fine results. As there are no alternative non-optimal direct global methods to provide the initial model parameters, the workaround is to exploit

- Factorization methods,

- Sequential methods,

- Global motion first methods.

The factorization methods are apt for projective reconstructions and were first to provide global initial approximations for large image blocks. They seem no longer to be sustained because of the big downside that all points must be visible in all views [Sturm and Triggs 1996]. Sequential methods are the workhorse of the majority of existing bundle adjustment software packages. These category of methods encompass

- the classical merge of independent models which finalizes with a single absolute orientation [Kraus 2007],

- starting from a minimal solution (2,3 views), the cycle of resection, intersection and refined BA takes place where the methods might differ in the philosophy of incrementing subsequent images to the global image block (e.g. hierarchical processing) [Fusiello 2014; Snavely et al. 2008].

All sequential approaches are at risk of error accumulation and the high sensitivity of the goodness of selecting the first minimal solution. Recently, a new class of multiple view reconstruction methods handling unordered sets of images has emerged. The global motion first methods convey a completely different logic towards image block orientation and aim to fight the weakness of current methodologies. Like in the other cases, the approach is 3-step, starting at a local scale of 2 or 3 views (recovery of the so-called *relative motions*), to finalize with a single merge of the *frames* (rather than points) in the global system (*global motions*) and the BA refinement.

The general pipeline for the global reconstruction process consists of the following steps

- obtain matches and pairwise rotations,

- rotation estimation,

- translation recovery,

- bundle adjustment and dense 3D recovery.

First off, the epipolar graph is built and the pairwise rotations R_{ij} are factored out from the essential matrices. Then, all pairwise rotations are placed in a common matrix G which is solved for the global rotations R_i, R_j. In particular, the maximal trace is searched for

$$max \sum_{i,j=1}^{n} trace\,(R_{ij}^T R_i^T R_j^T) = max_G\, trace(G^T G). \tag{2.25}$$

To find the global translation vectors, Arie-Nachimson et al. [2012] reformulate the essential matrix equation translation (see Equation (2.26) and compare with Equation (2.24)), solving stack of linear equations (see Equation (2.27)) for every pair of corresponding points (in local system). Nonetheless, one can find deviations from this formulation, e.g. using a trifocal sensor instead. Lastly, improved camera rotations, translations, and the global structure is calculated in the BA.

$$E_{ij} = R_i^T(\mathbf{T_i} - \mathbf{T_j})R_j \tag{2.26}$$

$$\mathbf{p}_i^T R_i^T(\mathbf{T_i} - \mathbf{T_j})R_j \mathbf{p}_j = 0 \tag{2.27}$$

The novel approach is said to be much less demanding in terms of computational powers, and less sensitive to outliers. The processing is "light" because, unlike in the sequential techniques, here the number of unknowns depends solely on the number of input images. The global structure is computed, if desired, only in the latest stage. It is of no use in the steps responsible for the motion (camera orientation) computation. More importantly, there is no need to provide the initial pair to build the global orientation. See further references for deeper insights [Arie-Nachimson et al. 2012; Arrigonis et al. 2014; Martinec and Pajdla 2007; Moulon et al. 2013].

2.4 Specular surfaces

Light at specular interfaces does not maintain the collinearity as in perspective imaging, but is piecewise linear [Kutulakos and Steger 2008]. Section 2.1 explains the physical causes of the phenomena. Successive paragraphs tell what basic geometric and photometric properties are inherent in the particular setting; as well as how the properties can deed shape information. It is summed up with a concise review of state-of-the-art measurement techniques, making the distinction between the photogrammetric and computer vision approaches.

2.4.1 Geometric properties

Moving specularities. Diffuse surfaces by definition reflect equally in all directions, nonetheless the everyday surfaces also typically exhibit specular components. As observed at the start of the chapter, smooth surfaces do not posses an appearance of their own, but reproject what is around them. The projection itself is determined by surface shape, illumination- and observer-positions, according to which the incident power is asymmetrically concentrated onto a distinct direction. Blake and Brelstaff [1988] point out that the notion of disparities in specular environments is ill-defined (and not following the epipolar constraint), for what is observed is a relative change with respect to physical features; see Figure 2.10. Albeit the specificity, specular highlights do carry
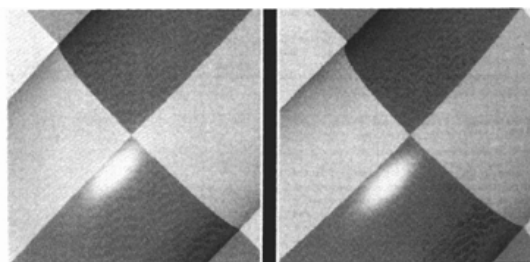

Figure 2.10: An example of a specularity whose "motion" relative to surface features is oblique, i.e. vertical relative disparity is not zero [Blake and Brelstaff 1988].

information about local surface shape. A qualitative check on convexity/concavity can be performed by moving the observer in front of and behind a spoon. If the experiment shows the highlights move with the observer, its shape is convex, otherwise it is concave. Another qualitative behaviour is that specularities travel fastest in places of small curvature and are therefore more commonly seen on highly undulating shapes [Zisserman et al. 1989].

Apart from this, quantitative measures can be drawn from the moving specularities by stereoscopic viewing, known as *specular stereo* and sketched in Figure 2.11. The mandatory knowns are the relative position $\mathbf{d}$ of left and right cameras, position of the light source and a reference point $\mathbf{C}$ in vicinity of the occurring specularity $\mathbf{A}$. As thoroughly covered in Blake [1985], the depth $\mathbf{V}$ and the surface normal $\mathbf{N}$ can be computed one after another by noting three cycles

$$\mathbf{V} + \mathbf{d} - \mathbf{r} - \mathbf{W} = \mathbf{0}, \tag{2.28}$$

$$\mathbf{V} - \mathbf{L} - \mathbf{S} = \mathbf{0}, \tag{2.29}$$

$$\mathbf{L} - \mathbf{L'} - \mathbf{r} = \mathbf{0}, \tag{2.30}$$

29

writing down the law of reflection as

$$\mathbf{N} = \frac{\mathbf{V} + \mathbf{L}}{|\mathbf{V} + \mathbf{L}|}, \tag{2.31}$$

$$\mathbf{N}' = \frac{\mathbf{W} + \mathbf{L'}}{|\mathbf{W} + \mathbf{L'}|}, \tag{2.32}$$

and imposing the reference point to C to lie close enough to the point of occurring specularity A that both are contained within the same tangent plane

$$(\mathbf{F} - \mathbf{V}).\mathbf{N} = 0. \tag{2.33}$$
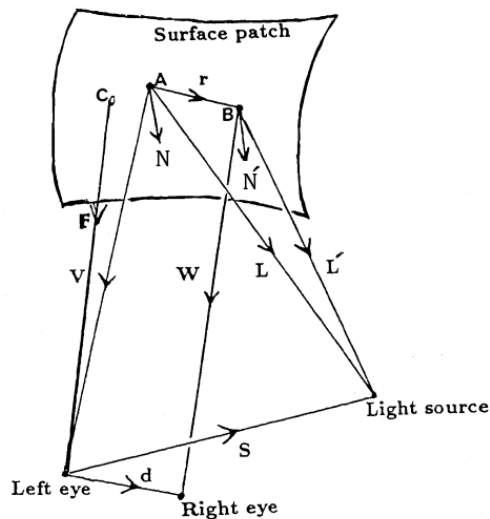
Figure 2.11: Geometrical arrangement of the specular stereo. A smooth surface illuminated by a point light source along **L** [Blake 1985].

Depth-normal ambiguity. In a more general case, there is only one camera and either the reference point **C** (see Figure 2.11) is missing, or the illumination source cannot be regarded at infinite distance from the surface[9]. The incidence direction then becomes dependent on the surface element height, apart from the source position, as Sanderson et al. [1988] reports. The authors carry on to say that on such occasions one needs to deal with the depth-normal ambiguity, i.e. for a specular highlight reprojected to image coordinates there exists a family of normal vectors θ and surface heights z that

---

[9]Besides the distant source, the surface must be assumed to occupy a fixed height and span an area that is small compared to object-source radius.

satisfy the imaging condition. As reads Figure 2.12 (left) and Equation (2.34), when the x coordinate can be fixed by the camera, and the source is placed at $S(x_s, z_s)$, all elements reflecting the light lie on a parabola, with a being a parameter for the parabola family [Sanderson et al. 1988]

$$(x - x_s)^2 = 4a(z - z_s + a).$$ (2.34)

The same principle holds for refractive surfaces, if the refractive index is known [Morris 2011]. A pattern placed below the water surface and recorded on the other side of the medium does not provide enough information to know where the viewing ray intersects the water interface; see Figure 2.12 (right).
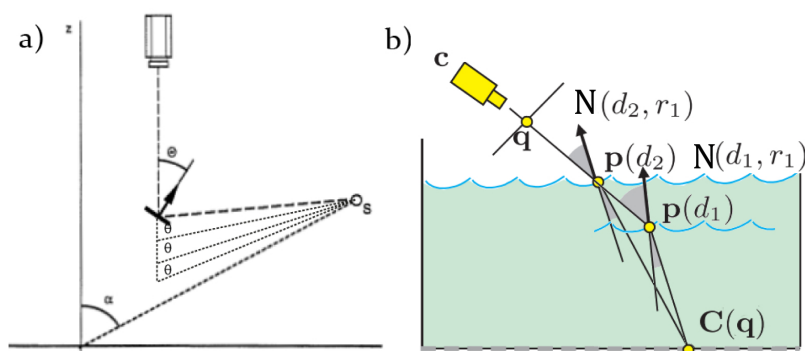

Figure 2.12: Depth-normal ambiguity on specular surface (left) and within refractive media (right). Adapted from [Morris 2011; Sanderson et al. 1988].

A second view can remove the ambiguity, but the light source position or the underwater pattern must act as knowns. A thorough mathematical background to the problem is given in Morris [2011]; Sanderson et al. [1988]; Savarese and Perona [2001].

Direct ray measurement. To remove the normal-depth ambiguity, incidence rays can be restored through direct measurements. It requires identification of at least two points on the incident ray, and is irrespective of the type of surface [Kutulakos and Steger 2008]. Triangulating the $\vec{KL}$ ray with the ray $\vec{OI}$, which originates at the camera's perspective centre and intersects the image plane, returns the depth of points on the object of interest; see Figure 2.13. The normal in that point can be calculated as the angle bisector of the incident and exitant directions.

In the field, finding two reference points that project to a single pixel calls for a controlled environment. One way to achieve this is to move a calibrated pattern around the measurement space in a manner that allows for the association of 2+ reference points to each or only selected sensor elements.
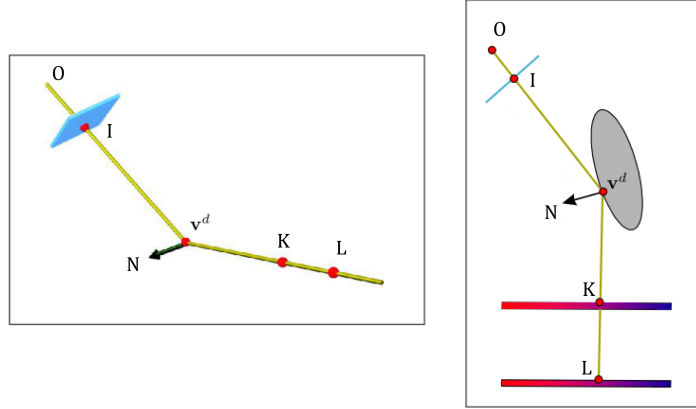
Figure 2.13: Identifying two known scene points allows for aliminating the depth-normal ambiguity with a path triangulation method, in transparent (left) and specular (right) conditions. Adapted from [Kutulakos and Steger 2008].

2.4.2   Photometric properties

Rather than tracing the light rays and using the law of reflection, one can recourse to mathematical modelling of specular reflection, including the creation and use of reflectance map. Two common specular models are described here — the *simple specular model* and the *Torrance-Sparrow model*. The Torrance-Sparrow is also the foundation of another commonly deployed model — the *di-chromatic specular model* [Shafer 1985]. The latter is not investigated hereafter. The photometric properties are discussed only in the context of specular surfaces.

The reflectance map and the simple specular model.   Ikeuchi [1981] revised the photometric stereo (see Paragraph 2.3.1) to point out its suitability to manage specular surfaces. In order to build the reflectance map for a specular surface, three relationships between the source radiance and the perceived image intensity are established, i.e. (i) between the light source radiance and the incident radiance in Equation (2.35), (ii) between incidence radiance and emitting radiance in Equation (2.36), (iii) between emitting radiance and the image irradiance (perceived intensity) in Equation (2.37).

$$L_i = L_s \tag{2.35}$$

$$L_r(\theta_r, \phi_r) = L_i(\theta_r, \phi_r + \pi) \tag{2.36}$$

$$I = \{(\frac{\pi}{4})(\frac{d}{f})^2 \cos^4 \alpha\} L_r \tag{2.37}$$

where I, d, f, α are the image irradiance, the camera diameter of the entrance aperture, focal length and the off-axis angle, respectively. From the last equation it is apparent that the intensity one perceives in the image is proportional to the surface radiance, while the factor of proportionality is a composition of the imaging system specific constants. From Equation (2.35) it is derived that the incidence radiance is equal to the source radiance, and Equation (2.36) states that for a specular surface, no matter how complicated the source distribution is, merely a single direction contributes to the the exitant radiance. This conclusion is meaningful as it implies that pixel intensity is dictated by the orientation of the surface patch, yet the brightness of that patch is equal to the brightness of the part of the light source that it happens to reflect at a particular instance, that is

$$R(p_n, q_n) = I(p_n, q_n) = \gamma L_r(p_n, q_n) = \gamma L_s(p_s, q_s). \tag{2.38}$$

The Torrance-Sparrow specular model. Healey and Binford [1988] laid down a methodology to calculate local orientations, principal curvature and direction from image brightness measured on a specularity. They showed that, unlike most contemporaneous approaches which were founded on the very general reflection geometry, experiments suggest more complex reflectance models are required. Sticking to the simplest specular model, specular points would be observable in single points. This is untrue for the majority of specular surfaces will glare around the reflection patch. Moreover, specularities shall manifest only in cases when vectors $\mathbf{V}, \mathbf{N}, \mathbf{L}$ lie in the same plane, which corresponds to $\alpha = 0$ in Figure 2.14. Again, it is invalid if one assumes a non-ideally reflecting surface, for which the underlying roughness will cause a reflecting cone, rather than a reflecting single line towards the viewer or the camera.
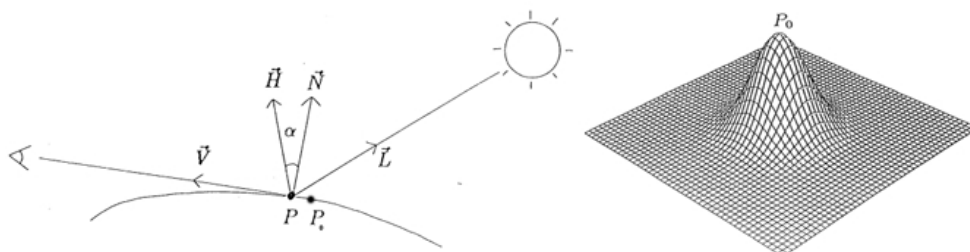
Figure 2.14: Left: Geometry of a specular reflection. Right: Specular intensity surface. The grid lines have constant specular intensity. Adapted from Healey and Binford [1988].

The authors adopted the Torrance-Sparrow model in their investigations, defined as a multiplication of the Fresnel coefficient F, the facet orientation distribution function D, and the geometrical attenuation factor A (see Equation (2.39)). The Fresnel coefficient depends on the incidence angle, and the refractive index of the reflecting material. The facet distribution function is characteristic of the Sparrow-Torrance model which models the surface as an assembly of small micro facets. The D function for the Gaussian

distribution is presented in Equation (2.40), where K is a constant, m describes the surface roughness, and for a given α the value of the function is proportional to the fraction of facets oriented towards $\mathbf{H}$. Having eliminated from the specular model those parameters that can be regarded constant for an analysed setting, see Equation (2.41).

$$I = FDA. \tag{2.39}$$

$$D = Ke^{-(\alpha/m)^2} \tag{2.40}$$

$$I = K'e^{-(\alpha/m)^2} \tag{2.41}$$

So long one deals with a surface of Gaussian curvature, spotting its point of maximum intensity is trivial. This very point signifies that $\mathbf{N}$ aligns with $\mathbf{H}$, and lets the immediate calculation of the orientation at that spot (see Figure 2.14). With the known intensity I' at $\mathbf{P_0}$, the orientation at an arbitrary $\mathbf{P}$ can be devised by reorganizing Equation (2.41) to

$$|\alpha| = m\sqrt{-\ln\left(\frac{I'}{K'}\right)}. \tag{2.42}$$

Computing the α value for the entire specular intensity image gives rise to the α image. From the α angle image one can infer the local surface shape and so describe it up to the second order. Refer to the original publication for complete derivations.

2.4.3 In Photogrammetry

The handling of untypical surfaces has been a core challenge of photogrammetry for many years and has been solved using both passive and active techniques. In aerial applications, Tewinkel [1963] analysed a number of factors possibly influencing the perception of the water depth and proposed a method to correct for the refraction bend in restituted depths. Accounting for the multi-media bending effect was also treated by Höhle [1971], who laid down the analytical and analogue methods for restitution in underwater environment. Since then, many successful projects were conducted using the principles of underwater photogrammetry. These approaches vary according to the object in focus. By and large they are dominated by rigorous modelling of the media [Harvey and Shortis 1995; Maas 1995; Maas et al. 1993; Menna et al. 2013; Mulsow 2010; Shortis et al. 2013], or by conversion to an opaque appearance [Chandler et al. 2014; Cobelli et al. 2009a; Han and Endreny 2014], for which well established measurement methods exist. Exploiting their inherent characteristics is seldom, motivated by the fact that complex methods which base on refraction of specular reflection become unmanageable for reproduction in the real world ambient.

The same holds true for glossy, specular or transparent solid objects where straightforward solutions take the lead; coating, or attaching physical targetting enables direct measurements [Blandino et al. 2004; Hinz et al. 2009].

One can also employ active sensing by taking into account the material properties and illuminating it with the part of the light spectrum, which is reflected in a quasi Lambertian way. For instance, the infrared band will work for water and the UV band for transparent foils [Hilsenstein 2004; Rantoson et al. 2010]. Another example was introduced by a group of scientists at *NASA* who developed an active method for inspecting membrane sails used in the Space. They dope a fluorescent dye within the material structure. The dye is invisible to a human eye but when excited with a laser of appropriate wavelength, it returns a very strong signal back to the laser [Dorrington et al. 2010; Jones et al. 2003]. Thanks to this property, the camera equipped with a selected cut-off filter becomes capable of distinguishing see-through and mirroring objects from the surrounding scene

Satisfactory results are also achievable with traditional multiview approaches. Because specularities move with the observer, it is relatively easy to eliminate them from the measurement. If one has a calibrated three-camera rig, then a point/patch in one view (specular reflections inclusive) is projected with the fundamental matrix to remaining views and if some similarity measure is inconsistent, the point/patch is likely to correspond to a specular reflection [Bhat and Nayar 1998]. Other methods exist for getting rid of specular reflection from a scene, such as polarization or color filtering [Nayar et al. 1991; Park and Kak 2004; Wells and Danehy 2005]. On the other hand, if the highlights should be of interest, one can identify correponsing specularities in stereo or multiview imaging systems with the help of the Helmoholz principle. This category of method is often referred to as fusion of photometric stereo with multi-view stereo. An example of such a system is presented in [Kiefhaber 2014].

All the approaches mentioned utilize passive sensors in reconstructing the surface shape. Active illumination techniques applied to highly reflecting surfaces produce systematic errors in depth estimation. There are few recorded attempts to adapt the scanning techniques to deal with the specularity effects, in particular by space-time analyses of the signal and filtering procedures [Park and Kak 2008].

2.4.4 In Computer Vision

The way in which the computer vision community confronts difficult surfaces is much more avantgarde. Rather than finding a way around the problem, as is often the case for photogrammetrists, they intend to face the problem directly. These inventions, however, seem rarely to leave the strictly theoretical or laboratory conditions.

The two principal classes of algorithms are *shape from distortion* and *shape from specularity*. Among them, there are approaches operating both in a calibrated and an uncalibrated environment. The calibrated environment may imply a calibrated point, line or a pattern whose reflections are registered by a camera setup. In the following paragraphs, the most important methods of each of the algorithm classes will be recalled. The polarization based techniques are not reviewed hereafter. For more detailed explanation of the methodologies the reader is referred to Ihrke et al. [2008]; Stolz et al. [2012] or the individual publications cited in the text. Further interesting approaches that fall outside the scope of the adopted categories exploit other (than the visible) parts

of the electromagnetic spectrum, such as infra-red [Meriaudeau et al. 2010], thermal [Eren et al. 2009] or UV. The principle resembles that of Hilsenstein [2004]; Rantoson et al. [2010], already discussed in the preceding subsection. One searches for a wave spectrum, in which the information/signal received from the problematic surfaces — either doped or hit by the energy portion — is maximised, while minimising the close-by, disturbing signals.

Shape from distortion Because specular scene recovery from a single image is generally ill-posed, the researchers always introduce various constraints such as planarity, smoothness or distant observer and illumination into their algorithm pipelines. To start with, an interesting biomedicine application was presented by Halstead et al. [1996]. The authors project a series of concentric circles onto a human cornea and capture with a camera their deformed state. At the same time, given some rough assumption of the cornea shape, they generate a synthetic image of the pattern. The guessed surface is then iteratively improved by normals fitting.

Bonfort and Sturm [2003] incorporate a calibrated pattern into their 3D scene and observe it from multiple calibrated views. The scene in front of every camera is discretized into voxels along the projection rays and all voxels are assigned a surface normal vector. The main methods of this approach traverse all the scene voxels to reject the least credible ones and keep the ones consistent across the views and the calibration pattern. The curve fitting work of Kutulakos and Seitz [2000] is in the same tone.

Savarese and Perona [2001] demonstrated that the local shape of a curved mirror surface can be restituted if an image of two known intersecting lines are recorded by a camera. A second order surface was shown to be retrievable from the orientation and curvature of two intersecting lines, whereas the third order description necessitated an extra third line. Oren and Nayar [1997] made the distinction between real and virtual scene features. As the observer or camera moves, the real features will remain on the surface they physically belong to, yet features that are reflections of the ambient scene will change their positions. Based on that, and bringing in the notion of caustic reflections the authors could detect the virtual features and track their trajectories as the camera moved. Then, some derived differential equations related the feature trajectory with the observed surface profile. 2D, 3D cases are investigated and curves of arbitrary shapes are considered too.

A novel view on the specular scene called the *specular flow* was recently provided by several researches [Adato et al. 2007; Roth and Black 2006; Sankaranarayanan et al. 2010]. Roth and Black [2006] are said to be the first to introduce the concept of specular flow and they define it as follows. If a camera sees a specular surface that reflects its surrounding, then the motion of either the camera or the scene will give rise to 2D image motion of the specular reflections. The image motion is the *optical flow* [Horn and Schunck 1981] induced by some relative motion but since it is now determined in the specular context, it is denoted the specular flow. The advantage of specular flow with respect to other distortion-based methods is that no prior knowledge on the scene illumination is necessary other than it is far-field. The environment, in terms of photometric

properties can be arbitrary. The downside is, in analogy to the optical flow, in the enforced infinitesimal motion step. The surface in question is then related to the observed specular flow and known camera or environment motion field through the shape from specular flow equation. The equation is made up of nonlinear partial differential equations which makes it a rather non-trivial solution. However, the state-of-the-art shows that under new parametrization schemes and initial conditions, the solving difficulty can be greatly minimized and certain assumptions relaxed [Sankaranarayanan et al. 2010].

Shape from specularity The theory behind shape from specularity is not so much different from the one provided in the previous paragraph. Here, the specular highlights are exclusively in focus, and they are projections of calibrated illumination sources. The method of Ikeuchi [1981] and the method of Healey and Binford [1988] discussed in Subsection 2.4.2 are very early examples that fall within the group category.

Sanderson et al. [1988] developed a highlights scanning technique to obtain 3D surface information. An array of point sources is used to scan all possible positions and directions of incident light rays relative to the camera. It is equivalent to discrete scanning of the reflectance map (see Paragraph 2.3.1). A stereo camera is employed to disambiguate the depth to normal relation (see Paragraph 2.4.1).

Hertzmann and Seitz [2005] introduce the orientation consistency cue. Their hypothesis is that two points with the same surface orientation produce identical appearance in image space. The normals for each point on the target surface are guessed by finding their matches in other, reference objects placed in the observed scene. To robustify the method, the transfer of normals from the reference to the target surface can be performed a number of times under different illumination conditions. The shape is recovered by integrating over the normal vectors. The BRDF, the illumination and the camera are assumed to be unknown and arbitrary.

Some researches use the BRDF in recovering detailed information of the surface structure. Lensch et al. [2003] invented an image-based technique to recover Spatially-Varying BRDF without making assumptions about the material type, and even allowing heterogeneous material types. Having found the reflectance functions they also provide the surface orientations and a normal map. Chen et al. [2006] use an isotropic BRDF and images to recover surface mesostructure. Also in Wang and Dana [2006] the fine scale surface geometry was recovered. Just like in the former method, they measure the BRDF with an imaging system device and subsequently draw the normal map. The invention is in using a multi-views approach for the simultaneous capture of the surface reflectance and geometry.

The chief aims of the work of Ma et al. [2007] are twofold. First in deriving a shading algorithm to perform rendering of complex surfaces and secondly in creating a quality 3D model of surface geometry. They built a scanning system that employs spherical gradient illumination and diverse polarization patterns to recover surface normals. The first permits a lighting pattern that is independent of the view, while thanks to the latter the normals can be separately estimated for diffuse and specular components of the surface. Due to the fact that specular features alter the polarization direction of the

reflected light, while diffuse features return unpolarized signal, the polarizes prove very useful tools in telling them apart.

Direct ray measurement. The very recent and representative methods for direct ray measurement were developed in Kutulakos and Steger [2008]. They conceived a general framework to reconstruct shape of refractive and specular surfaces objects. In contrast to what has been presented so far, this method allows the light to pass through or be reflected from up to two different media and surfaces, respectively. The way from 3D points to 3D rays to 2D image points is naturally piecewise linear. The objective of the work was to identify a set of surfaces that can be reconstructed given solely the information about vertices along the light path. This led to an algorithm that can find the mirror surface observed with one camera provided that light path bounces only once, and there can be two points identified along its trajectory (see Figure 2.13). Moreover, it was demonstrated that for unknown scenes where two bounces occur, at least three cameras are necessary to carry out the light path triangulation.

3

Parameter estimation

3D shape reconstruction of the previous chapter is possible only with known camera orientation parameters and image observations. These are linked in the bundle adjustment via the collinearity equations or other constraints. Recovering the cameras and the structure (3D scene) is thus essentially a parameter estimation problem. This chapter will review some of the most common estimation techniques applied in vision tasks, i.e. those based on multi-views and correspondences. It will be shown that estimating parameters in the bundle adjustment proposed by photogrammetry, and computer vision, as well as Kalman filtering, derive from the same theorems and are equivalent under certain conditions. Börlin and Grussenmeyer 2013; Förstner and Wrobel 2006; Kraus 1997; Maybeck 1982; Mikhail and Ackermann 1976; Nocedal and Wright 2006; Triggs et al. 2000 were used as the main references for this chapter. See Förstner 2013 for a very interesting proof of Bayesian networks and conditional random fields being a generalized bundle adjustment parameter estimation problem.

3.1 Least squares methods

Estimating parameters is a task present across many, if not all, disciplines of science. Given a set of measurements, one may want to fit a model with its parameters into the data, and reduce the problem to few parameters that enable reasoning on the observed phenomena in a simplistic manner. One may as well demand data at times (prediction) or in places (interpolation) where the original measurements do not deliver information. Or, as the measurements are never exact but subject to random errors, giving results that are different up to the probabilistic fluctuations, the objective may be to recover the unique and most optimal estimates of observations. Finally, the interest is in the goodness of fit of the data into the model, as well as the accuracy of estimating the model's parameters. In photogrammetry, geodesy and computer vision it is typically carried out through statistical modelling where the variables are regarded as stochastic,

and their estimates as well as distributions are obtained in a least squares adjustment (LSA) procedure.

The mathematical model The adjustment with least squares method is defined by a *mathematical model* which abstracts some physical situation. It further divides into *functional* and *stochastic models*. The former is a conceptual, deterministic representation of the problem to which the measurements correspond, hence it relates the observations with the unknown parameters (alternatively merely relating the parameters). The latter describes nondeterministic, i.e. probabilistic, properties of the variables involved. For instance, if the functional model is the collinearity equation that relates the image space with the object space, then the variability of image observations, camera parameters as well as the designation of free and fixed parameters constitute the stochastic model.

Least squares principle Because the redundant measurements do not yield a unique result despite complying with the given model, there is a need to include an additional criterion. When the original set of observations is denoted as l, and their estimates adjusted based on the employed mathematical mode are $\hat{\mathrm{l}}$ (Equation 3.1), the known *least squares principle* (Equation 3.2) states that the vector of residuals **v** shall guarantee that the output $\hat{\mathrm{l}}$ is as close as possible to the observed l. In Equation 3.2 the weight matrix $\{W$ reflects stochastic properties of the observations. Its diagonal elements correspond to precision, whereas the off-diagonal elements describe correlation present among observations. The formulation does not require any assumption on the distribution of observations.

$$\mathbf{v} = \hat{\mathrm{l}} - \mathrm{l} \tag{3.1}$$

$$\Phi = \mathbf{v}^T W \mathbf{v} \rightarrow minimum \tag{3.2}$$

Observations and parameters Besides observations, other variables that may participate in the LSA are the parameters (unknowns). They are functionally associated with the measurements and likewise posses the stochastic characteristics. If treated as unknowns, they will be freely adjusted over the course of computations; if treated as observed unknowns, their observed values will be refined as the adjustment progresses yet their change will be constrained by the precisions defined in the weight matrix. Lastly, although unrealistic, some parameters might be seen as constants, that are free of error thus with fixed values in the adjustment. In practice there are no parameters free of error, and inclusion of constants with stochastics or observed unknowns is recommended [Kraus 1997] (*unified least squares adjustment* as coined by Mikhail and Ackermann 1976). Another motivation to abstain from *constants* is due to the fact that the error propagation is bypassed and resulting *a posteriori* precision estimates (standard deviations derived from covariance matrix) are rendered too optimistic. On the other hand, in photogrammetry, if the cameras and 3D points are well arranged, in

other words good imaging network is ascertained, fixing some parameters that are inherently stochastic does not underestimate the final covariance matrix. In literature it is recognized as *Total Error Propagation* and *Limiting Error Propagation* [Fraser 1987].

Conditions and constraints   Four variants of relationships between observations $\mathbf{l}$ and parameters $\mathbf{x}$ can be distinguished,   (i) with observations and no parameters in Equation 3.3, (ii) with one observation and no parameters in Equation 3.4, (iii) with many observations and parameters in Equation 3.5, and (iv) with parameters only, in Equation 3.6.

$$f(\hat{\mathbf{l}}) = \mathbf{0} \tag{3.3}$$

$$\hat{\mathbf{l}} = f(\hat{\mathbf{x}}) = \bar{\mathbf{l}} + \mathbf{v} \tag{3.4}$$

$$f(\hat{\mathbf{x}}, \hat{\mathbf{l}}) = \mathbf{0} \tag{3.5}$$

$$f(\hat{\mathbf{x}}) = \mathbf{0} \tag{3.6}$$

Equations including the observations are termed *condition equations* and are formulated with explicit or implicit functions (e.g. collinearity or coplanarity equations). In photogrammetry, the explicit formulation is handled by the Gauss-Markov model, while the implicit formulation is handled by the more general Gauss-Helmert. By contrast, functions of pure parameters are always implicit and known as the *constraints*. They can be estimated with both Gauss-Markov and Gauss-Helmert models. In Section 3.1.1 an in-depth treatment of adjustment w/o constraints using the Gauss-Markov model is given.

Numerical optimisation techniques   Having selected a suitable mathematical model, the remaining part is the choice of a computational algorithm. Two contrasting least squares techniques are used by the photogrammetric and computer vision practitioners respectively. The "photogrammetric way" has been based on the Gauss-Markov theorem, stating that for a linear system, the least squares weighted by the true measurement uncertainties (W in Equation 3.2) give the *Best Linear Unbiased Estimator* (BLUE), i.e. it gives the minimum variance. The non-linear functional model is linearised to the *Linear Substitute Model* and solved with the Newton's method. Finally, improved approximates of the unknowns are then found in an iterative manner.

A somewhat different philosophy comes from the fields dealing with vision tasks but rooted in computer science. It is motivated by the need to efficiently and automatically handle big and difficult datasets contaminated with considerable numbers of outliers. There, the least squares are solved with non-linear optimisation techniques where the aim is to find the minimiser of an objective function that is linearised to a quadratic local model (see Equation 3.23). Preferred optimisation techniques are the Gauss-Newton (line search) and Levenberq-Marquardt (trust-region). The principal difference between the two approaches and the Newton's method in the Gauss-Markov model is in (i) the local approximation of the functions, and (ii) how the parameter updates are applied.

Over the course of recent years there has been an observable trend to employ the methods relevant for computer science also in photogrammetry. More exhaustive discussion will follow in Sections 3.1.1 and 3.1.2.

3.1.1 In Photogrammetry

Collinearity equations are non-linear, form the basis for all the adjustment processes performed within photogrammetry and constitute the functional model of the LSA. Linearisation of the equations by Taylor expansion is a standard and allows the adjustment to be treated with the linear least squares method. The optimal results are obtained in a sequence of iterative runs of linear LSA.

Linear Gauss-Markov model

Given the type of relationship between observations and unknown parameters as in Equation 3.4, where $\bar{\mathbf{l}}$ is the vector of observations, $\mathbf{v}$ is the vector of estimated residuals and $\hat{\mathbf{x}}$ are the estimated parameters, one wants to find the corrections to approximate parameter values

$$\Delta\hat{\mathbf{x}} = \hat{\mathbf{x}} - \hat{\mathbf{x}}^0 \tag{3.7}$$

$$\Delta\hat{\mathbf{l}} = \mathbf{l} - f(\hat{\mathbf{x}}^0), \tag{3.8}$$

and subsequently solve the *Linear Substitute Model*

$$\Delta\hat{\mathbf{l}} = \Delta\mathbf{l} + \hat{\mathbf{v}} = A\Delta\hat{\mathbf{x}}, \tag{3.9}$$

where A is the Jacobian after u parameters in n observation equations

$$A_{n,u} = \frac{\partial f(\mathbf{x})}{\partial\mathbf{x}}, \tag{3.10}$$

which should optimally be re-linearised at each iteration step, i.e. for each new approximation of the parameters.

The stochastics are embedded in the weight coefficient matrix C_{ll} (aka covariance matrix), calculated from the weight matrix as

$$C_{ll} = \sigma_0^2 W_{ll}^{-1}. \tag{3.11}$$

The σ_0 value is the reference variance and can be *a priori* (σ_0) or *a posteriori*

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^T C_{ll}^{-1}\mathbf{v}}{n - u}. \tag{3.12}$$

The *a priori* estimate is given prior to the adjustment and, provided the weight matrix correctly models the uncertainty of observations, it equals 1. The *a posteriori* estimate is computed from observational residuals and *a priori* weight matrix of the observations. So long as no aggressive robust techniques are incorporated within the solution, the two

figures intertwined play a crucial role in assessing the quality of the adjustment, and may serve as a criterion to terminate the adjustment iteration process.

Finding the best estimates of the parameters involves minimising the quadratic form in Equation 3.2, i.e. setting its first derivative to zero and having a second derivative vanish by default following the linearity assumption. It leads to the following system of normal equations,

$$A^T C_{ll}^{-1} A \Delta \hat{\mathbf{x}} = A^T C_{ll}^{-1} \Delta \mathbf{l}, \tag{3.13}$$

estimated covariance matrix,

$$C_{xx} = (A^T C_{ll}^{-1} A)^{-1}, \tag{3.14}$$

observational residuals,

$$\hat{\mathbf{v}} = A \Delta \hat{\mathbf{x}} - \Delta \mathbf{l}, \tag{3.15}$$

and their covariance matrix,

$$C_{vv} = C_{ll} - A^T C_{xx} A^T. \tag{3.16}$$

For large image blocks, the system of normal equations in Equation 3.13 requires large computational effort and storage capacity. In practice, the sparse matrices are effectively decomposed into factors with *Cholesky, LU, QR* or a *Minimum Degree Algorithm*.

Having completed an iteration, the *a posteriori* reference variance ($\hat{\sigma}_0$) ought to be compared against its *a priori* value (σ_0). If the difference is significant, it might indicate an incorrectly defined functional model. Controlling the rate of change as the iterations go, is one way to decide whether the adjustment has converged and can be stopped. If the solution has not been reached, the current estimates of unknowns are updated, and the cycle re-starts with Equations 3.8, 3.10, 3.13, 3.15 and 3.12.

Linear Gauss-Markov model with constraints between unknowns

Given condition equations (see Equation 3.4), and constraints between parameters (see Equation 3.6), the additional dependencies are enforced with the method of constrained minima by Lagrange multipliers, and the least squares criterion in Equation 3.2 evolves to

$$\Phi = \mathbf{v}^T W \mathbf{v} + 2\mathbf{k}(H\Delta\hat{\mathbf{x}} + \mathbf{w}) \rightarrow minimum \tag{3.17}$$

where H is the Jacobian after u parameters in h constraint equations, and the vector $\mathbf{w}$ contains the differences between the function value evaluated at current approximate values $\hat{\mathbf{x}}_0$ and their ideal values $\mathbf{0}$. The vector is non-stochastic and depends solely on the currently approximated values of parameters.

The Linear Substitute Model is hence extended by h constraint equations,

$$H\Delta\hat{\mathbf{x}} + \mathbf{w} = \mathbf{0} \tag{3.18}$$

$$H_{h,u} = \frac{\partial h(\mathbf{x})}{\partial \mathbf{x}}, \tag{3.19}$$

while the parameters get the extra Lagrange multipliers $\mathbf{k}$. The extended normal equations present as follows:

$$\begin{bmatrix} A^T C_{ll}^{-1} A & H^T \\ H & 0 \end{bmatrix} \begin{bmatrix} \Delta \hat{\mathbf{x}} \\ \mathbf{k} \end{bmatrix} = \begin{bmatrix} A^T C_{ll}^{-1} \Delta \mathbf{l} \\ -\mathbf{w} \end{bmatrix}. \tag{3.20}$$

Solving the system with the Cholesky decomposition is impossible because of the 0s on the matrix diagonal that violate the condition of semi-positivity. Gauss elimination can be used alternatively.

The estimated covariance of the parameters takes the following form:

$$C_{xx} = \begin{bmatrix} A^T C_{ll}^{-1} A & H^T \\ H & 0 \end{bmatrix}^{-1} \tag{3.21}$$

and the vector of observational residuals $\mathbf{v}$ with its covariance C_{vv} execute as in Equations 3.15 and 3.16. The *a posteriori* reference variance is given by

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^T C_{ll}^{-1} \mathbf{v}}{n + h - u}. \tag{3.22}$$

3.1.2 In Computer Vision

Here, the bundle adjustment is also defined in terms of conditions, constraints and the quality metric that one wants to minimize (see Equation 3.2). The parametrization, however, is different. 3D points are expressed in projective geometry, and rotations in the algebraic quaternions (see Section 2.2). The parameter space remains highly non-linear and requires appropriate solving techniques. The non-linear cost model is by standard expanded into a local quadratic model with the Taylor series

$$\Phi(x + \Delta x) \approx \Phi(x) + \mathbf{g}\Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}H\Delta\mathbf{x} \tag{3.23}$$

where $\mathbf{g}$ is the gradient vector, H is the Hessian matrix and J is the Jacobian of the model

$$\mathbf{g} \equiv \frac{\partial \Phi}{\partial x} = \mathbf{v}^T W J$$
$$H \equiv \frac{\partial^2 \Phi}{\partial x^2} = J^T W J + \sum_i (\mathbf{v}^T W)_i \frac{\partial^2 \mathbf{v}_i}{\partial \mathbf{x}^2} \tag{3.24}$$

The clear asymmetry regarding the techniques considered in the previous Section is the presence of the Hessian matrix in the expansion. As long as in photogrammetry the collinearity equations are linearised for all parameters and as such they land in the cost function, here the problem contained in the cost function is regarded non-linear and approximation up to the second order is carried out.

This quadratic expression can be solved with the Newton's method, which assures quick convergence but has a high cost per iteration and is susceptible to converging to a saddle point i.e. a non-global minimum. These predicaments are overcome by upgrading the method to *Gauss-Newton* or *Levenberg-Marquardt,* which are considered in the coming sections.

44

Gauss-Newton

Minimizing the non-linear problem defined in Equation 3.23 with the Newton method is performed by setting the first derivative to zero i.e. $H\Delta\mathbf{x} = -\mathbf{g}$ which eventually gives the *Newton step* in the form of

$$\Delta\mathbf{x} = -H^{-1}\mathbf{g}.$$ (3.25)

Yet, because the last term of of the Hessian is likely to be small relative to the first term, it is plausible to approximate the Hessian only with the first term and save both the implementation trouble and the computational time. Especially the latter, due to the Jacobian that constitutes the first part of the Hessian and is already computed for the gradient vector. This gives rise to the *Gauss-Newton* method and normal equations

$$J^T W J \Delta\mathbf{x} = -J^T W \mathbf{v}.$$ (3.26)

The derived normal equations are identical to the normal equations derived from linear LS in Equation 3.13, and the $\Delta\mathbf{x}$ is indeed the solution to Equation 3.9. However, unlike in the linear scenario, the $\Delta\mathbf{x}$ in the non-linear framework is not regarded the ultimate correction to current approximates of the parameters, but merely as the descent direction (as long as the Jacobian is full-rank). The length of the step α along the search direction

$$\hat{\mathbf{x}} = \mathbf{x}^0 + \alpha\Delta\mathbf{x}$$ (3.27)

that will minimize the objective function is still to be found. One of the two accepted conditions to arrive at the α value are *Wolfe* and *Armijo conditions*.

Levenberg-Marquadt

The *Levenberg-Marquadt* method (aka *Damped Newton* method) can be seen as an upgrade of the Gauss-Newton method. It replaces the line search with the trust-region strategy, and overcomes the disadvantage of the Gauss-Newton, allowing for the handling of rank-deficient Jacobians. For a spherical trust-region, the algorithm works by restricting the magnitude of the allowable descent to a radius r. After performing an iteration with the Gauss-Newton method, the value of the descent is verified, and if it falls inside the region defined by r, it concurrently become its solution. Otherwise a λ parameter (see Equation 3.28) is introduced along the diagonal of the normal equations. The parameter ascertains that $\Delta\mathbf{x} = \Delta\mathbf{x}_{LM}$ satisfies $||\Delta\mathbf{x}|| = r$, and if $\lambda \to \infty$ then the solution becomes the naïve gradient descent with $\Delta\mathbf{x} \propto -\mathbf{g}$.

$$(J^T W J + \lambda I)\Delta\mathbf{x} = -J^T W \mathbf{v}$$ (3.28)

3.2   Recursive estimation

Recursive estimation is defined as a probabilistic approach for estimating an unknown probability density function recursively over time. The need for recursive estimation was born owing to the facts that (i) no perfect and absolute mathematical model to describe

real processes exists, and (ii) there are no measurement devices that can provide perfect data. Hence, given the knowledge about a signal with its current state and some new, heterogeneous data provided by a noisy measurement, the goal is to estimate the true state of the signal conditioned on a selected optimality criterion. For instance, the state might correspond to the position of an aeroplane at time t_i, while the measurements are a combination of e.g. the speed provided by an airspeed indicator, the GNSS and IMU data. Knowing the dynamics of the system and statistical description of the present noises, the task of recursive estimation is then to (i) predict the aeroplane's position at time t_{i+1} given both the state plus the estimates of necessary parameters at time t_i, and (ii) combine this prediction plus its uncertainty with the measurements performed at time t_{t+1} also including its uncertainties.

Kalman filter

The *Kalman filter* (KF), which will be considered next, is (i) an example of recursive estimation, (ii) acts as a data interpolator ($t < t_1$), data filter ($t = t_1$) or data predictor ($t > t_1$), (iii) builds upon the mathematical foundation of probabilistic theory and least squares, and (iv) was invented to tackle the limitation of the *Wiener filter*.

Initially, to confront the unclear mathematics of the Wiener filter, Kalman approaches the problem from the point of view of conditional distribution and expectation. No other statistics than first (mean) and second order (variance) moments are required within the processing. Thanks to this property, KF can process the data less computationally expensive manner. Random signals are then said to be the output of an exclusively linear dynamic system excited by independent random signals, i.e. white noise. By introducing the concept of the *state* and *state transition*, and assuming the linearity of the system, Kalman resolves the Wiener issue of growing memory filters. There is now only a single derivation, a first-order difference equation, that ties two adjacent time steps and is sufficient to predict. The *state* is said to be a set of numbers that characterize a dynamic system, the dynamics themselves, i.e. how the state passes from one state to another, are enclosed inside the *state transition*. Expressed in mathematical terms, the transition of a linear system from $\mathbf{x}_t$ to $\mathbf{x}_{t+1}$ is a differential equation of the form

$$\mathbf{x}_t = F\mathbf{x}_{t-1} + G_{t-1}\mathbf{w}_{t-1}$$
$$P(\mathbf{w}) \sim (0, Q),$$
(3.29)

while the measurements $\mathbf{z}_t$, thus the system output, being functionally related to the state, are as follows

$$\mathbf{z}_t = H_t\mathbf{x}_t + \mathbf{v}_t$$
$$P(\mathbf{v}) \sim (0, R).$$
(3.30)

The F in Equation 3.29 is often called the transition matrix, and the entire equation is known as the *time update*. On the other hand, the H of the subsequent equation is the measurement matrix and the entire expression is called the *measurement update*. In the adopted derivation the $\mathbf{v}, \mathbf{w}$ are assumed independent random Gaussian variables, with

means at $\mathbf{0}$ and covariances R, Q, respectively. In the above formulations and further digression no control input is assumed to be present.

There is more than one way to infer the KF equations and not all require assumptions on the variable distribution. Yet, as the aim of KF is to find the best estimate of the state $\mathbf{x}$ at some instance in time, conditioned on the knowledge of the data coming from the measurements $\mathbf{z}$, the apparent conditional tie brings the Bayesian viewpoint into attention. The Bayesian estimator (see Equation 3.31) is the *a posteriori* density of the parameter vector $\mathbf{x}$ for an observed vector $\mathbf{z}$ explicitly given as a product of the joint density $f(\mathbf{x}, \mathbf{z})$, and the prior $f(\mathbf{z})$ that in the case of KF would be the measurement prediction.

$$f(\mathbf{x}|\mathbf{z}) = \frac{f(\mathbf{x}, \mathbf{z})}{f(\mathbf{z})} \tag{3.31}$$

Embarking on the Bayesian interpretation collates the KF with the *Hidden Markov Models* (HMM) (see Figure 3.1) and the Gauss-Markov model (already discussed in Section 3.1). It is a *Markov process* as it fulfils the *Markov property*, i.e. the conditional probability distribution of the predicted state depends only upon the current state (indeed $\mathbf{w}$ is a white process), and not the states that proceeded it. It is a HMM because it is Markov and because the true states are not directly visible but elucidated from the device output (compare Equation 3.30). Lastly, the assumption of the Gaussian random variable and the Markov property make the KF a Gauss-Markov model.
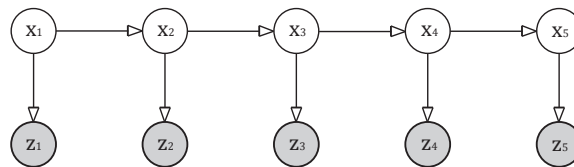

Figure 3.1: Kalman filter without control as hidden Markov chain. The nodes contain the x and z that are the unobserved system states, and the independent measurements upon which the state depends, respectively. The directed edges specify conditional probabilities.

Kalman Filter forward reasoning

The filter operates in cycles of forward reasoning, that is of repeating

- predictions, where given the current state estimate and the estimate of the covariance, KF propagates them to the next time instance, and

- updates, where it integrates the prediction with the new measurement to generate the best state estimate; the fusion is done via Bayesian estimation using the prediction as a prior.

$$\hat{\mathbf{x}} = argmax_x f(\mathbf{x}|\mathbf{z}) = argmax_x \frac{f(\mathbf{x}, \mathbf{z})}{f(\mathbf{z})}. \tag{3.32}$$

If the initial state of the system is unknown, it is modelled as a random vector that is normally distributed around the anticipated mean, taking into account the covariance matrix. Derivation of the KF equations will follow in the successive paragraph.

The conditional density function in Equation 3.31 can be rewritten to a form in Equation 3.33, the conditional mean $\mu_{x|z}$ (first order moment) of which is a function of x, as rendered in Equation 3.34. Accordingly, the conditional covariance $P_{x|z}$ (second order moment) presented in Equation 3.35. μ_x, P_x and μ_z, P_z are the joint density means and covariances of the state and the measurement, respectively.

$$f(\mathbf{x}|\mathbf{z}) = \frac{1}{(2\pi)^{\frac{\pi}{2}}|P_{x|z}|^{\frac{1}{2}}} exp\{-\frac{1}{2}[\mathbf{x} - \mu_{x|z}]^T P_{x|z}^{-1}[\mathbf{x} - \mu_{x|z}]\} \tag{3.33}$$

$$E_{\mathbf{x}}[\mathbf{x}|\mathbf{z} = z] \triangleq \mu_{x|z} = \mu_x + P_x P_z^{-1}(\mathbf{z} - \mu_z) \tag{3.34}$$

$$E_{\mathbf{x}}\{[\mathbf{x} - \mu_{x|z}][\mathbf{x} - \mu_{x|z}]^T|\mathbf{z} = z\} \triangleq P_{x|z} = P_x - P_{x,z} P_z^{-1} P_{x,z} \tag{3.35}$$

Let's presume that the predicted estimates $(\mathbf{x}_t^-, P_t^-)$ are known, yet the $(\mathbf{x}_t^+, P_t^+)$ and $(\mathbf{x}_{t+1}^-, P_{t+1}^-)$ are to be devised. Up front, the mean and covariance of joint density will be deduced, followed by the recovery of moments of the conditional density function. The t subscript will be omitted in the primal phase, that is no subscript signifies the implicit time t. By rewriting the KF measurement step as shown in Equation 3.36 and transferring the result into Equations 3.34, 3.35, one obtains the mean (see Equation 3.37) and the covariance (see Equation 3.38) of the joint density.

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{z} \end{bmatrix} = \begin{bmatrix} \mathbf{x} \\ H\mathbf{x} + \mathbf{v} \end{bmatrix} = \begin{bmatrix} I & 0 \\ H & I \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{v} \end{bmatrix} \tag{3.36}$$

$$\mu_{x,z} = \begin{bmatrix} I & 0 \\ H & I \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_t^- \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{x}}^- \\ H\hat{\mathbf{x}}^- \end{bmatrix} \triangleq C\mu \tag{3.37}$$

$$P_{x,z} = \begin{bmatrix} I & 0 \\ H & I \end{bmatrix} \begin{bmatrix} P^- & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} I & H^T \\ 0 & I \end{bmatrix} = \begin{bmatrix} P^- & P^- H^T \\ H P^- & H P^- H^T + R \end{bmatrix} \triangleq C \sum C^T \tag{3.38}$$

At that point, substituting the immediate results with Equations 3.34, 3.35 gives way to the conditional mean $\hat{\mathbf{x}}^+$ and covariance P^+ in Equations 3.39, 3.40. These equations constitute the eventual form of the measurement update, and appear frequently in a reduced form, including the *Kalman gain* (see Equation 3.41). Note that the updated state estimate is an outcome of the predicted state plus a weighted difference between real measurement and its predicted version, where Kalman gain stands for the weighting factor. To close the cycle, the next step is to run the time update. The current optimal state and its covariance are propagated to the future instance (see Equation 3.42, 3.43).

48

Changing the filter structure with the help of Q would indicate increased or decreased trust in the adequacy of the underlying model. By similar reasoning, increased R suggests measurements are subject to elevated noise levels. Furthermore, due to conditional covariance P being statistically related to the measurement history through H^- and R^- and at the same time not being the functions of measurements, it can be precomputed prior to the measurements. A similar rule applies to the Kalman Gain K

Important to stress is the fact that whether the KF is launched in a recursive manner (single matrix inverse per time instance) or in a batch (all time instances in a single equation system), the output yields equivalent results. This is of tremendous relevance in online processing tasks where prompt answers are anticipated.

$$\hat{\mathbf{x}}^+ = E_{\mathbf{x}}[\mathbf{x}|\mathbf{z} = z] = \hat{\mathbf{x}}^- + [P^- H^T][HP^- H^T + R]^{-1}[z - H\hat{\mathbf{x}}^-] \tag{3.39}$$

$$P^+ = P^- - [P^- H^T][HP^- H^T + R]^{-1}[HP^-] \tag{3.40}$$

$$\mathbf{K} = P^- H^T[HP^- H^T + R]^{-1} \tag{3.41}$$

$$\hat{\mathbf{x}}_t^- \triangleq E\{\mathbf{x}_t|\mathbf{z}_{t-1} = z_{t-1}\} = F\hat{\mathbf{x}}_{t-1}^+ \tag{3.42}$$

$$P_t^- \triangleq E\{[\mathbf{x}_t - \hat{\mathbf{x}}_t^-][\mathbf{x}_t - \hat{\mathbf{x}}_t^-]^T|\mathbf{z}_{i-1} = z_{t-1}\} = FP_{t-1}^+ F^T + GQ_{t-1}G^T \tag{3.43}$$

Kalman Filter and least squares methods

One may bring the updated step equation to an algebraically different form of Equations 3.44 and 3.45. Now, suppose that there is no prior information on the state i.e. $P^- = \infty$ or $(P^-)^{-1} = 0$, the update step reduces to Equations 3.46 and 3.47.

Note that if $W = R^{-1}$, the results are identical to those derived in the least squares solved with Gauss-Markov and Gauss-Newton methods in Section 3.1 (compare Equations 3.13 and 3.26). The difference is in the interpretation of the two matrices, namely, the least squares theory gives no insight into the nature of the measurement noise $\mathbf{v}$ hence W can be general as long as it reflects the measurement uncertainties. KF, on the contrary, imposes Gaussian noise.

$$\hat{\mathbf{x}}^+ = [P^+(P^-)^{-1}]\hat{\mathbf{x}}^- + [P^+ H^T R^{-1}]\mathbf{z} \tag{3.44}$$

$$P^+ = [(P^-)^{-1} + H^T R^{-1}H]^{-1}. \tag{3.45}$$

$$\hat{\mathbf{x}}^+ = P^+ H^T R^{-1}\mathbf{z} = [H^T R^{-1}H]^{-1}H^T R^{-1}\mathbf{z} \tag{3.46}$$

$$P^+ = [H^T R^{-1}H]^{-1} \tag{3.47}$$

System calibration and triangulation

This chapter opens up the experimental part of the dissertation. After a concise review of the related work on acquisition systems, and camera calibration, the adopted methods will be delved into. The content of the chapter covers the findings described in Rupnik and Jansa [2013, 2014].

4.1 Related work

Acquisition systems

The trend that brings photogrammetry close to novel and individual approaches is clearly reflected in the range of publications covering many domains. Typical examples are: close-range mapping with unmanned aerial vehicles (UAV) [Nex and Remondino 2014; Schneider et al. 2013], mobile mapping [Van Den Heuvel et al. 2006], recording of cultural heritage [El-Hakim et al. 2007; Remondino and Campana 2014; Remondino 2011], human motion analyses and a long array of industrial applications. For instance, development and testing in aerospace industry [Pappa et al. 2002, 2006; Shortis and Johnston 1996] , quality inspections in automotive manufacturing and renewable technologies [Bösemann 1996; Mostofi et al. 2012], in ship industry for reverse engineering [Menna and Troisi 2010] and in construction [Lin et al. 2008] for online quality control, robot guidance, as-built monitoring surveys, or material testing [Maas and Hampel 2006].

The spectrum of applications is wide, and so is the spectrum of approaches. As a general rule, *on-line* systems, a.k.a. *real-time*, are preferred to observe dynamic events, and when immediate results are expected. Employed sensors include smart or machine vision cameras. According to Maas [2008], the latter are defined as cameras accompanied by a host computer whereto the data-streams are directly written, e.g. Proscilia GE,

PCO Dimax, GOM ARGUS 5M, AICON MoveInspect HR. On the contrary, smart cameras are stand-alone devices (as well as dSLRs), integrate on-chip processing units and often return only dimensional coordinates rather than raw images, e.g. Optotrak Certus, Qualisys Oqus. In either case, the market offers a good selection of on-line systems in terms of varying spatial and temporal resolutions. The technology is mature and automated to the degree that no expert knowledge is necessary to operate it. Unfortunately, the prices are correspondingly higher in comparison to systems that will be discussed in the following paragraph, while the accuracies worse due to limited redundancy.

Off-line systems are the appropriate choice when the scene is static or almost-static, with the rate of change smaller than that of subsequent image acquisitions, and the immediate results are not required e.g. monitoring of a dam, reconstruction of a cultural monument. The principal tool of off-line systems are professional stand-alone digital single-lens reflex (dSLR) cameras (e.g.Nikon D3x, Canon EOS-1Ds), available from the consumer market [Bösemann 2011]. dSLR cameras are valued for flexibility and reasonable price but because the devices are not inherently built for metric purposes, they lack mechanical stability, be it the fixing of the sensor plane w.r.t. the housing of the camera. Still, recognizing the caveats, understanding their physical cause and consequence is the key to a successful i.e. high precision, measurement. The great ally of off-line applications is the time. It allows for careful survey planning and capturing the scene with a favourable network of images thus ensuring fine point distribution, decent intersections, and recovery of instantaneous camera calibration parameters. When combined with coded targets, the reconstruction process can be reduced to a few mouse clicks. Consequently the need for repetitive retrieval of camera interior orientation through self-calibrating is nowadays viewed as a routine background task rather than an additional effort [Fraser 2013].

Bösemann [2011]; Luhmann [2010]; Maas [2008] unanimously claim dSLRs not to be the right devices for dynamic observations. It is a fair conclusion considering the above limitations, and the fact that direct interfacing is impossible. Nonetheless, few scientists have struggled to prove it empirically. Most of the reported cases use consumer grade cameras in multi-exposure acquisitions at frequencies less than 1Hz [Benning et al. 2004; Detchev et al. 2013; Koschitzki et al. 2011] . Much lesser interest is found in applying the cameras at higher frame rates, that is substituting single images for videos [Nocerino et al. 2011].

Camera calibration

The camera calibration describes the camera frustum by (i) the principal point, and (ii) principal distance (IOR); as well as lens- and sensor-related errors. In Equation (2.10) they are denoted as x_0, y_0, c and $\Delta x, \Delta y$, respectively. The latter two are often called *systematic image errors* or *distortion corrections* and they correspond to the non-linear perturbations present in the images. Lens distortions are usually grouped into *radial* and *decentring,* caused by variations in refraction indices across the lens surface, and the uncentered lens elements along the optical axis, respectively. The sensor errors are

characterized by *in-plane* and *out-of-plane* distortions caused by shrinkage/stretching of film cameras and sensor unflatness.

Considering the techniques that allow the retrieval of bespoken parameters, one can single out (i) *laboratory*, (ii) *test field*, (iii) *on-the-job* and (iv) *self-calibrations*. The laboratory calibration, as the name suggests, is done in a lab environment. Nowadays, only high precision aerial cameras still undergo that procedure. The test field calibration involves a planar or spatial calibration rig of known dimensions. The on-the-job calibration is done simultaneously with the actual measurement, or shortly before it. It uses previously measured control points or alternatively a spatial rig of known dimensions as the reference for computing camera calibration in a self adjusting bundle. Last but not least, the self-calibration can be seen as the on-the-job or test field calibration except no control information is required.

To model the non-linear errors one can take the physical or phenomenological view. That is, one may try forth to correct for distortions caused by physical phenomena, otherwise the aim may be to simply reduce the image distortions, without the regard of their physical origin. In either case, one solves a curve fitting problem [Tang et al. 2012], and when approached with a polynomial function (which is true for the majority of the approved methods), the high correlations and over-parametrization issues must be addressed.

In close range photogrammetry, the physical viewpoint has been adopted throughout the last 40 years. The 8-parameter calibration model (three IOR, three radial and two decentring parameters) of Brown [Brown 1956, 1971] and its 10-parameter extension (with additional two in-plane distortion parameters) of Fraser [Fraser 1997] proved sufficient in both the analogue and digital era of imaging. Besides the Brown equations, also the Torlegard polynomials set has been commonly applied to model radial, decentring distortions as well as skewness and scaling of the axes [Torlegard 1967].

The larger the sensor area, the more the systematic image errors manifest. As a result, in the airborne domain, the phenomenological viewpoint is more often employed, and there exists no golden method to model the image errors. Yet, Ebner [1976], Grün [1978] and Jacobsen [1982] are among the most popular polynomial models. Recently, due to the new developments on the market, such as multi-head cameras, the topic of camera calibration is again in focus [Tang et al. 2012].

In computer vision, the term *auto-calibration* has been popularized. Among auto-calibration techniques, the noteworthy are those that exploit the *Kruppa* equations, and those that rely on constraints such as motion, plane or vanishing line constraints. Sensor misclosures are often called the *projective* or *affine distortions* and they describe linear, i.e. straight line preserving perturbations. The non-linear components of the distortion are less of a concern and if encountered then only up to the radial case [Zhang 2000].

For futher information the reader is referred to Clarke and Fryer [1998]; Hartley and Zisserman [2003]; Luhmann [2010]; McGlone [2004]; Tang [2013] which served as reference textbooks for this section.

4.2 Problem definition

The objective of the research was to develop a measurement system to determine the motion of a flexible platform[1] floating over the water surface. One of the challenges was in the fact that the equipment and the methods were being developed before the actual model was devised, let alone existing. The prior knowledge available at that time included the need to tackle

- non-rigid motions[2] of unknown rate and magnitude,

- difficult surfaces characterized by transparency and strong reflectivity,

- scalability as the capability to adjust to the varying object size and the workplace,

- high-precision, and

- low-cost.

Ultimately, the constructed platform composed of four identical sub-modules of size $1m \times 1m$, making it a $4m \times 4m$ model as a whole (see Figure 4.2). It hoovered above the water level thanks to a air-filled cushions, one for each module. As the platform was regarded flexible, its outerwear was approximated by a dense grid of retro-reflective targets to be registered by the cameras.

A series of measurements took place in a professional ship testing facility [*Vienna Model Basin Ltd*]. The motion was induced by water waves generated at regular and irregular intervals. The regular wave were of fixed amplitude, period and wavelength. During the measurement, the camera system was both static and undergoing a motion. An overview of the workplace situation is depicted in Figure 4.1.

The motion non-rigidity and the difficult surfaces were tackled by covering the surface with a dense grid of artificial retro-reflecting targeting. The low-budget solution and scalability was accomplished by employing consumer grade dSLR cameras, thereby, passive imaging. Passive photogrammetry — not requiring artificial illumination and not restricted to a finite volume of interest — allows to observe phenomena and objects at different scales by modifying the camera-to-object distance. Even though an illumination source is used throughout our acquisitions, it is only to enhance the signal-to-noise ratio, yet not needed for the smooth operation of the imaging system.

The accuracy of the results was confronted with a motion capture system (MCS)[3] that registered the model in parallel. It consisted of three Qualisys Oqus cameras [*QUALISYS Motion Capture Systems*], distributed on-site as shown in Figure 4.1. The adopted spatial resolution of the sensors was unknown, whereas the temporal resolution was fixed to 60 acquisitions per second. The system worked in the infra-red part of the spectrum and deployed spherical, retro-reflective targets.

---

[1]The *platform* will be interchangeably used with the *model*.

[2]Due to deformation effects, the platform motion was not reducible to a 6-parameter rigid trafo.

[3]In literature, motion capture systems are often abbreviated with MOCAP.

All the image processing tasks were carried out with the help of own developed software tools, while the image triangulation part (in static and dynamic cases) were handled by *Orient* [Kager 1989] — a photogrammetric adjustment package developed at the then Institute of Photogrammetry, Vienna University of Technology.

4.3 Adopted methods

4.3.1 Imaging system

In 2011 — at the time of conducting this research — as well as today, the full HD sensor is the highest resolution a state-of-the-art dSLR camera can offer. Over the course of the past year it has been observed that digital single lens mirrorless (dSLM) cameras have gained popularity and more increasingly offer video recording in the so-called 4K Ultra HD (UHD) technology, at up to 30fps; for instance: Sony DSC-RX100 IV, Panasonic Lumix DMC-G7, Nikon 1 J5.

The imaging setup comprised of three dSLR cameras (Canon 60D, 20mm focal length) and three continuous illumination sources (1250 Watt). Spatial resolution of the videos matched the full HD (1920x1080 pix), acquiring at maximum of 30 fps in progressive mode. The video files were lossy compressed with the H.264 codec and saved in a .mov container. It indicates that out of the 30 frames, only few are the so-called *key frames* hence entirely maintained. Others are obtained by sophisticated interpolation using the key frames and the block-wise differences kept by the encoder. Besides the motion picture compression, additional degradation in spatial resolution stems from the presence of the Bayer filter. The Bayer pattern is a 2x2 matrix that allows to derive colors — by interpolation — from the originally monochrome sensor. However, the author believes the loss of precision is much less aggravating than what is caused by the H.264 coding, and given that the targets were most of the time overexposed.

The mean object to camera distance amounted to 10m, resulting in the average image scale of 1:500. The cameras were rigidly mounted on a mobile bridge (see Figures 4.1, 4.4), and connected with each other, as well as with a PC, via USB cables to allow for (i) remote triggering, (ii) coarse synchronization. Fine-alignment of the video frames was possible with the help of a laser dot observed in all cameras. Despite the USB connections, the videos were stored on the memory cards. No spatial reference field was embedded in vicinity of the system, instead, the calibration and orientation was carried out with the *moved reference bar method*. Additionally, six scale bars were arranged along the model basin.

4.3.2 System calibration

By system calibration the authors refer to (i) interior orientation (IO) and lens distortion parameters (ADP), as well as (ii) relative orientation (RO) of cameras in a multi-ocular configuration (see Section 2.2). Two radial (3rd and 5th order) and two decentring distortion parameters of the Torlegard model were used to model the lens departures from
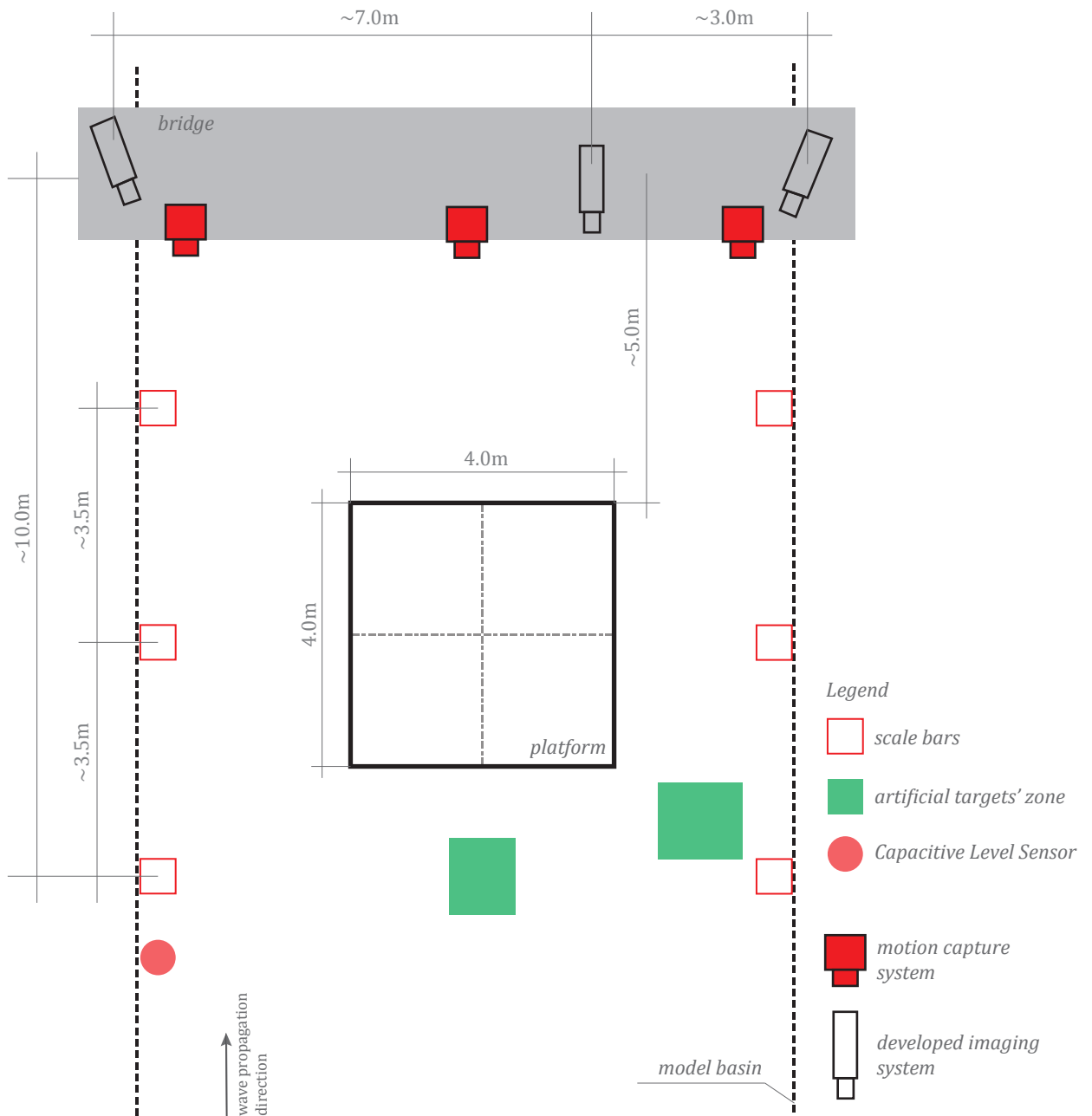
Figure 4.1: The top view of the measurement workplace. Cameras on top of the drawing are placed on a mobile bridge. The platform occupies the scene's central part, and is located ca. 5m away from the bridge. The red square markers indicate scale bars used in the *dynamic referencing* (see Section 4.3.4); the green rectilinear shapes correspond to the zones of the artifical targets (used in the *Method 1*, and in the evaluation of *Method 1* and *2*; see Subsection 5.3.2 and Section 5.4), while the red circle is the Capacitive Level Sensor (used in the evaluation of *Method 1* and *2*; see Section 5.4).

56
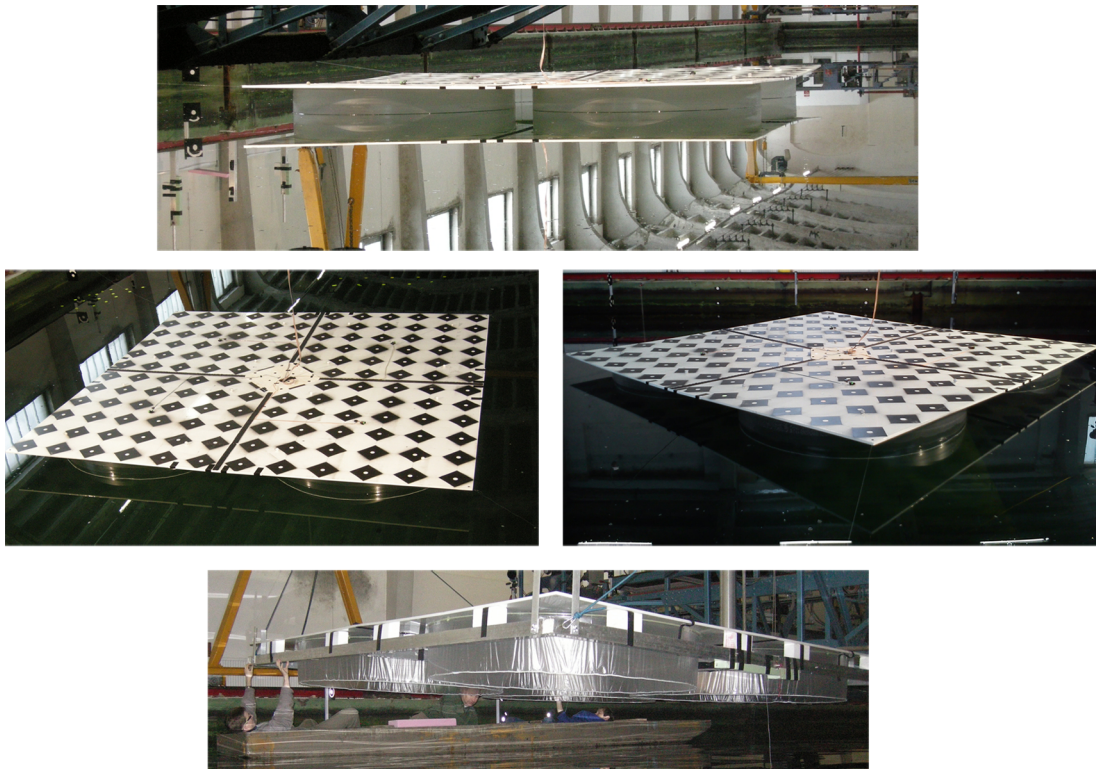
Figure 4.2: The platform placed on (top and middle images) and above (bottom image) the water. The top of the surface is discretized into a 12×12 grid of points. The suspension skirt is realized by four identical and symmetrically attached plastic foils. The top subfigure shows linear arrays being reflections of lamps hung above the basin.

Figure 4.3: The platform seen by the motion capture system. The bright spots are the spherical targets which are tracked when the motion occurs.

Figure 4.4: Red circles point to three cameras placed on a moving platform, across the model basin. The blue rectangles point to three scale bars. The remaining scale bars are symmetrically arrange on the opposite side of the basin.
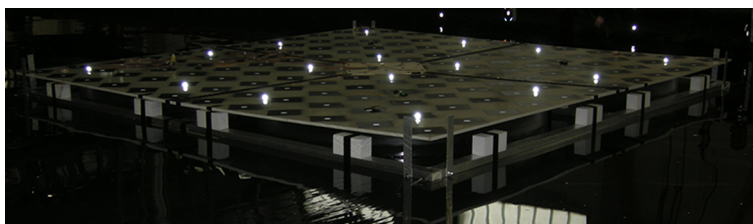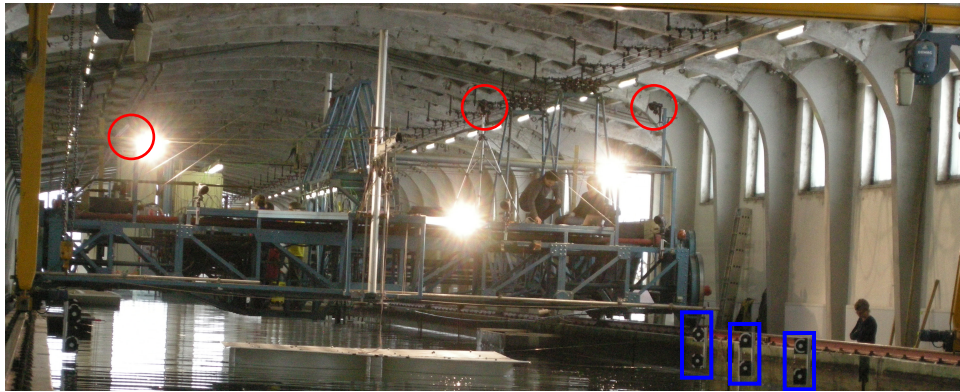
perspectivity. The cameras were used in previous surveys, thus regarded as precalibrated and included in the bundle adjustment as observed unknowns.

To assure the highest precision, thus reproduce the camera interior orientation and additional parameters at the instance of the measurement taking, the on site self-calibration was carried out. According to the photogrammetric network design rules, the self-calibration should employ varied camera roll angles, and a spatial field with well distributed points across the camera field-of-view, positioned at different depths in object space [Fraser 1997]. In our case neither of the conditions could be fulfilled. The cameras were fixed, and the object space, with the water occupying its large part, provided few points. Consequently, the method of *moved reference bar* was adopted.

The moved reference bar method uses a calibrated bar, signalised with targets on both ends (see Figure 4.5). The bar is randomly moved around the observation volume while the camera system tracks and records the positions of the two points in image space. Ultimately, the system calibration is calculated in the bundle adjustment, preferably with free network, including the image measurements and length of the bar as observations. The merit of the approach is that (i) interior parameters of all cameras and relative orientation are restored in one procedure, and (ii) the complexity of finding correspondences between particular views is significantly reduced [Luhmann et al. 2011; Maas 2008].

4.3.3 Tracking in image sequences

Point signalising in industrial applications commonly employ circular, retro-reflective (active) or white (passive) targets, surrounded by black, matte rings. Passive targets are more susceptible to ambient lighting and attention should be paid that there is enough of contrast to separate the points from the background. The advantage of retro-reflective

material is that when illuminated, it gives off a strong signal in the direction of the light source, which should also be the direction of the camera. Localizing the targets is then a trivial task as they are highly contrasted against the remaining image content. Once the points are localized, their centers are typically found with centroiding methods, ellipse fitting or cross-correlation [Burgess et al. 2011; Otepka 2004; Shortis et al. 1995; Wiora et al. 2004].

When tracking, physical environment often precludes a complete detection and localization of points. Firstly, freedom in applying targets and camera arrangement is hindered by workplace constraints. Secondly, as the observed object is dynamic, points may be occluded by other passing-by objects, or move beyond the field of view and be lost. Besides this, low-cost sensors are characterized by (i) lower resolution, and (ii) diminished image quality as a result of lower pixel sensitivity and applied compression. From the standpoint of mensuration algorithms, it translates into a loss of image measurement accuracy or even a loss of a tracked point.

Our experience showed that the latter is not a rare scenario. At a distance of 10m from the camera, with the decreasing imaging angle (but within the accepted angle range given by the retro-reflective sheet manufacturer), cross-correlation tracking was interrupted every few frames, whereas tracking based on thresholding techniques turned too unreliable in face of the poor targets' response. The situation could probably be amended if the lights were deposited closer to the object of interest, however, workplace constraints would not allow for that.

Figure 4.5: Left: the reference bar moved across the object space. Right: points moving away from the camera Line of Sight thus disappearing from the images.

To overcome the notorious loss of tracked points, we have modelled the motion of points in image space and implemented Kalman Filter (KF) (see Subsection 3.2) to detect their anomalous behaviour. The anomalous behaviour meaning (i) wrong centroiding due to glittering effects, (ii) loss of points due to temporary lack of illumination (points facing away the camera), or (iii) loss of points due to temporary occlusions (see Figure 4.5). Since we allow our image points to (i) manoeuvre freely, and (ii) accelerate/decelerate, the motion is described using the *acceleration model* [Mehrotra and Mahapatra 1997]. The state vector in Equation (4.1) consists of 2D image position, plus its first and second derivatives (velocity and acceleration). The transition and mea-

surement matrices are depicted in Equations (4.3) and (4.2). If the system kinematics are highly dynamic, one can extend this model further by including the third derivative (jerk).

$$\mathbf{X} = \begin{bmatrix} x & \dot{x} & \ddot{x} & y & \dot{y} & \ddot{y} \end{bmatrix}^T \tag{4.1}$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{4.2}$$

$$F = \begin{bmatrix} 1 & T & \frac{1}{2}T^2 & 0 & 0 & 0 \\ 0 & 1 & T & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & T & \frac{1}{2}T^2 \\ 0 & 0 & 0 & 0 & 1 & T \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{4.3}$$

where T is the duration of a single time frame. To use KF as a remedy for spotting bad data, we inspect the innovation values, that is the values predicted by the model. When the discrepancy between the model prediction and actual measurement is above a certain threshold, the measurement is either discarded, or replaced with the prediction for continuation purposes. Since KF works as an interpolator, we use it exclusively for detection of erroneous behaviour. We do not correct our measurements with the computed estimates unless qualified as erroneous.

The measurement workflow starts with the initialization phase, where the points of interest (circular retro-reflective targets) are detected and the correspondences in all views are established, followed by the tracking across time frames.

For the floating platform, the user draws a rectilinear shape along its borders, upon which a homography relating the image to object plane is determined. The points are then automatically found based on known platform grid, and measured with weighted centroiding. Provided that the relative orientation between cameras is given, the measurement is initialized in a single view and transferred to other views using fundamental matrices. Otherwise, initialization is performed for all existing views. Successive matching in time uses cross-correlation accompanied with 2nd degree polynomial fitting [Kraus 1997] for better precision.

The same procedure, apart from the homography-based definition of points of interest, applies to tracking of the reference bar or points placed on the water surface.

4.3.4 Dynamic referencing

At times the constancy of orientation elements is violated and cameras must be dynamically referenced to be able to align measurements from different epochs. Typical circumstances are (i) when the measurement volume must be enlarged during the measurement to compensate for the movement of the measured object, (ii) when the working environment conditions such as vibration affect the camera system position (image-variant interior orientation not reviewed hereafter).

60

Rotations and translations of the cameras are then considered unknowns, or observed unknowns. Their continuous restoration is possible when (i) a reference body is placed in the object scene, (ii) any static well-identifiable objects are present in the object scene, (iii) parameters' differences are observed by external sensors. The beauty of bundle adjustment permits then to combine all this information and output most optimal estimates of current camera positions.

When a reference body or external reference frame is available, transforming the measurements to the same datum can be computed via (i) 2D-3D resectioning in monocular measurements, or (ii) as sequential 3D-3D spatial similarity transformation in the multi-ocular case, and (iii) 2D-2D-...-3D simultaneous bundle adjustment. In either scenario the prerequisite is that the minimum of three static points can be observed [Bösemann 2011; Wrobel 2001].

Our three-camera measurement setup, placed on a bridge across the basin, observed a sequence of platform motions. To maintain the camera to object distance constant throughout the measurement, the bridge followed the platform (forward, backward motion). This change in camera pose directly affected the 3D triangulation result thus the task was to recover the motion of the camera setup and subsequently correct the 3D result. Given that (i) possibilities to include control information in proximity of water were limited, and (ii) the system of cameras was regarded rigid, we resolved the motion of cameras in a series of sequential 3D-3D orientations.

The adopted algorithm presents as follows

Data: image frames, GCP, RO, IO, ADP
Result: XYZ(t), EO(t)

for $t = 1$ **to** $frames$ **do**

    if $\mathrm{EO}^t \neq \mathrm{EO}^{t-1}$ **then**

        $xyz^t = \mathrm{BA}(\mathrm{EO}^{t-1}, xy^t)$
        $XYZ^t = \mathrm{BA}(\mathrm{EO}^t, \mathrm{GCP}, xyz^t)$

    else

        $XYZ^t = \mathrm{BA}(\mathrm{EO}^t, xy^i)$

    end

  end

Algorithm 1: Pseudocode of dynamic referencing. Marked in red are the unknowns at each step.

where xy are the image space coordinates, xyz are the object space coordinates before the motion correction (local frame), and XYZ are the coordinates after the motion correction (global frame), EO is the exterior orientation and BA is the bundle adjustment. System calibration parameters (RO, IO, ADP), control information (GCP) and images extracted from videos constitute the input to the algorithm. Whenever a change in control information in image space is recorded ($\mathrm{EO}^t \neq \mathrm{EO}^{t-1}$), points intersected at

this instance are regarded local (xyz), and are attributed to a 3D local coordinate system (EO^{t-1}), see Figure 4.6 (middle). The local CS is related to the global CS by a spatial similarity transformation (Equation (2.8) with $\lambda = 1$). To "subtract" the camera motion from the total motion, and obtain the 3D object points in the global frame (XYZ), least square adjustment is carried out. The unknowns in the adjustment are XYZ of all points and 6 parameters of the rigid transformation for an image frame. The basis of the transformation is the static, control information registered in both systems. Figure 4.7 illustrates the algorithm operation on real data.
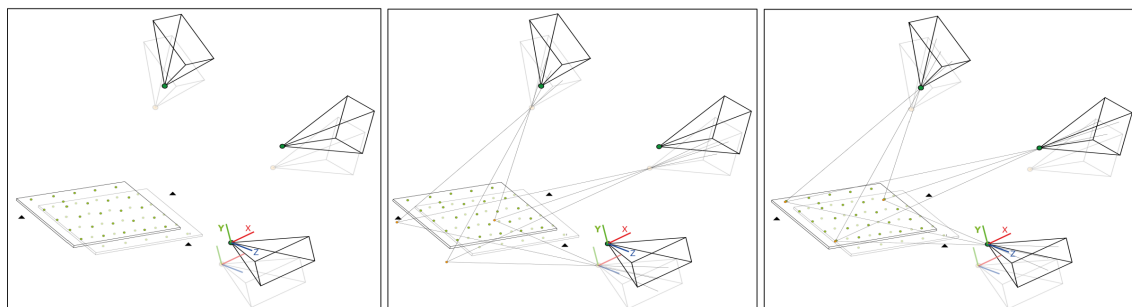

Figure 4.6: Dynamic referencing. Left: camera motion and object motion occured, middle: point reconstruction in the local camera frame, right: points after the spatial similarity camera motion correction.

4.4 Evaluation

In parallel to the operation of the developed system, the platform was observed by an online motion capture system (MCS) [*QUALISYS Motion Capture Systems*], of unknown but highly probably superior spatial and doubled temporal resolution. The two systems used independent targeting impeding their direct comparison. Unlike the flat and numerous (144 retro-reflective targets) targetting of our imaging system, the MCS employed few spherical points placed at an offset with respect to the true platform surface (see Figure 4.3).

In ship hydromechanics, analysing dynamics of bodies and fluids in regular waves is usually described in terms of *ship motions* (rendered in Figure 4.8). The motions can be split into three mutually perpendicular translations around a center of gravity (COG), and three rotations (*roll*, *pitch*, *yaw*) around respective axes. The axes convention is: X axis in the direction of wave propagation, and Z pointing upwards, perpendicular to still water surface [Journee and Pinkster 2002]. With this in mind, the accuracy of the results were validated (i) in regular wave conditions (static case) using the calculated ship motions (see Figures 4.11, 4.12, 4.13), and (ii) in the irregular spectrum (dynamic case) by comparing the reconstructed distances with their nominal values (see Figure 4.17).
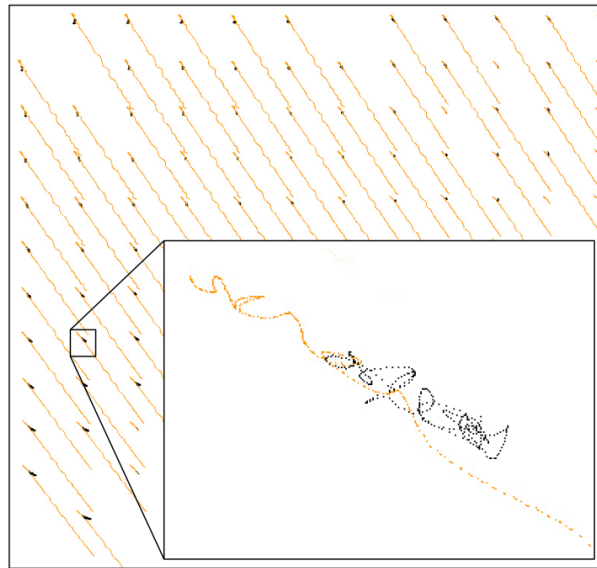
Figure 4.7: A set of reconstructed points (4D), in orange before and in black after the motion correction, cameras moving away from the object. The close-up figure: trajectory before motion subtraction overlaid with the corrected one; trajectories split up when cameras moved.

Static case

The floating platform consisted of four modules assembled together in a nonrigid fashion. The ship motions were thus independently calculated for each single piece since as such one may with higher probability speak of rigid manoeuvres.

The evaluation workflow is depicted in Figure 4.9. The reference coordinate system was defined parallel to the respective platform modules when in calm water conditions. It was indeed found by fitting a plane into the platform points, selecting the Z-coordinate to point upwards, and the origin to be in the centroid. For every analysed time instance, a new plane was fitted into the very same set of points, but in their new positions. By projecting the current normal vector to respective planes of the reference coordinate system, the rotational part of the ship motions was retrieved. For the translational part, the motion of the centroid of all points involved was used.

As a preprocessing step, the signals from our measurement system were upsampled to match the frequency of the MCS, and cross correlated for fine synchronisation. Figures 4.11, 4.12, 4.13 present the compared results in a graphical form, and the corresponding RMSE values are embedded in Figure 4.9.

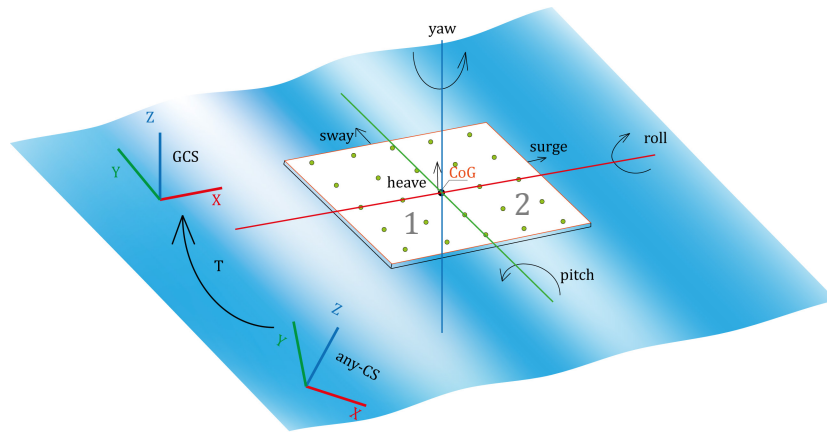**Figure 4.8:** Definition of the ship motions ($roll$, $pitch$, yaw). T denotes a rigid transformation. The enumerated modules (1 and 2) were validated against the MCS results.
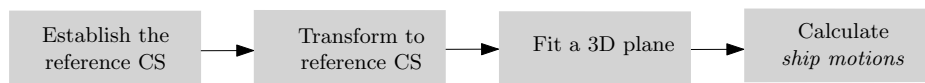

Figure 4.9: Evaluation workflow for the static camera case.

Series	Module	$\sigma_{x,y,z}$ [mm]	[1]$RMSE_{trans}$ [mm] surge/sway/heave	[2]$RMSE_{rot}$ [°] roll/pitch/yaw
1	1	1.8	2.6	0.07
		0.9	0.6	0.06
		1.1	1.8	0.03
	2	0.7	2.3	0.07
		0.5	0.5	0.05
		0.6	1.7	0.03
2	1	2.6	2.1	0.04
		0.9	0.7	0.07
		1.9	0.7	0.09
	2	2.1	1.4	0.09
		0.9	1.9	0.10
		1.9	1.2	0.10
3	1	1.3	1.0	0.04
		0.6	0.6	0.04
		1.0	0.7	0.05
	2	0.9	1.9	0.04
		0.5	1.1	0.06
		0.9	0.6	0.04

(Figure 4.11 spans Series 1; Figure 4.12 spans Series 2; Figure 4.13 spans Series 3)

Figure 4.10: The three static measurement series presented in figures. In each case two platform modules were analysed. *Module 1* and 2 were situated respectively further and closer to the cameras. $\sigma_{x,y,z}$[4] reports the output of the bundle adjustment (precision), whereas the $RMSE$ values result from the comparison against the MCS (accuracy).

---

[3] The global reference system was rotated such that the Y-coordinate ran along the wave propagation direction, and X across the basin. Here, for the sake of clarity of interpretation, the X, Y coordinates were swapped.

Figure 4.11: Measurement results (graphical presentation) of *Series 1* in terms of ship motions. The signal response in red comes from the developed measurement system, whereas the black refers to MCS. Surge in (a), (b), sway in (e), (f), heave in (i), (j), roll in (c), (d), pitch in (g), (h), and yaw in (k), (l). *Module 1* and 2 are defined in Figure 4.8.

Figure 4.12: Measurement results (graphical presentation) of *Series 2* in terms of ship motions. The signal response in red comes from the developed measurement system, whereas the black refers to MCS. Surge in (a), (b), sway in (e), (f), heave in (i), (j), roll in (c), (d), pitch in (g), (h), and yaw in (k), (l). *Module 1* and 2 are defined in Figure 4.8.
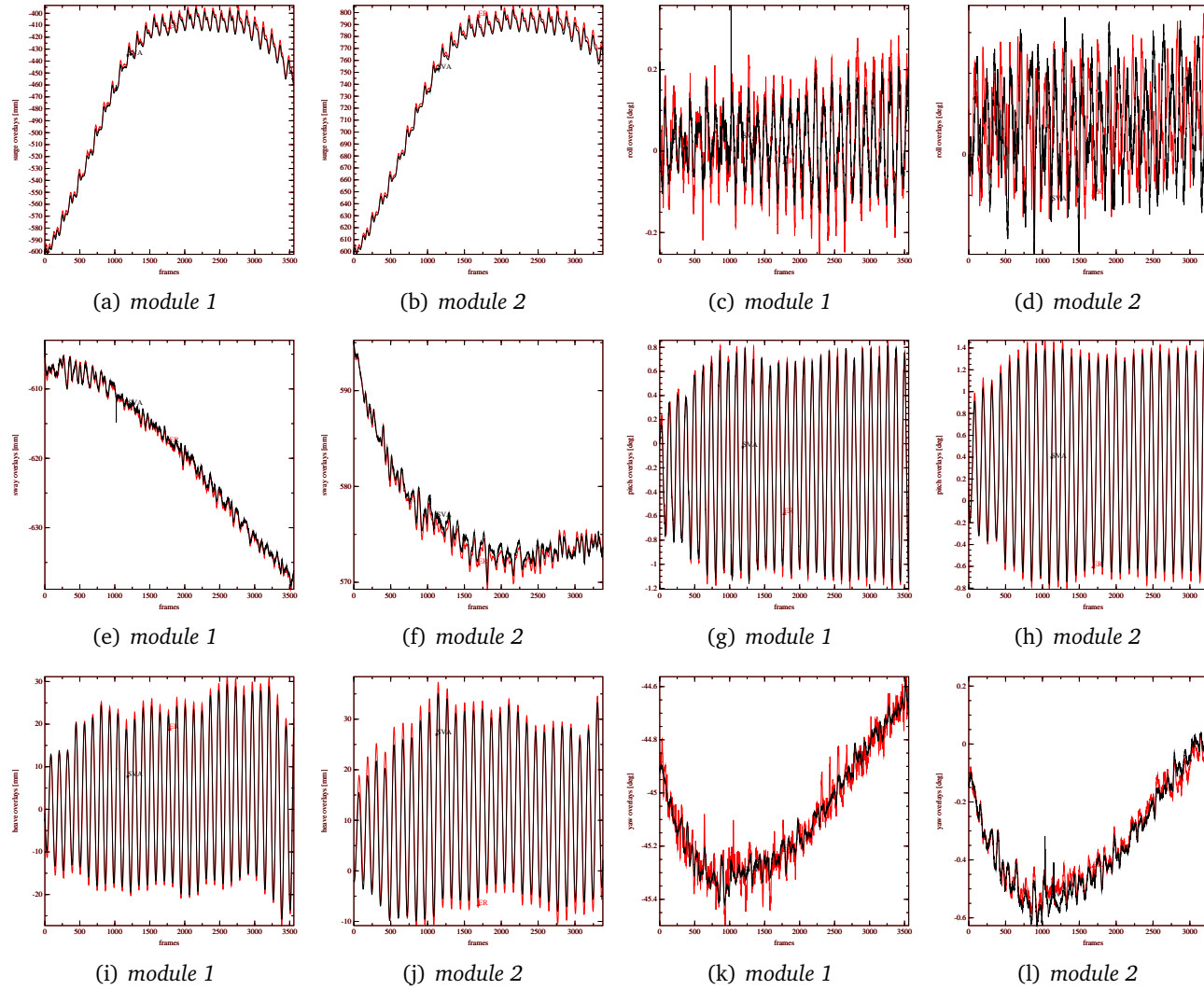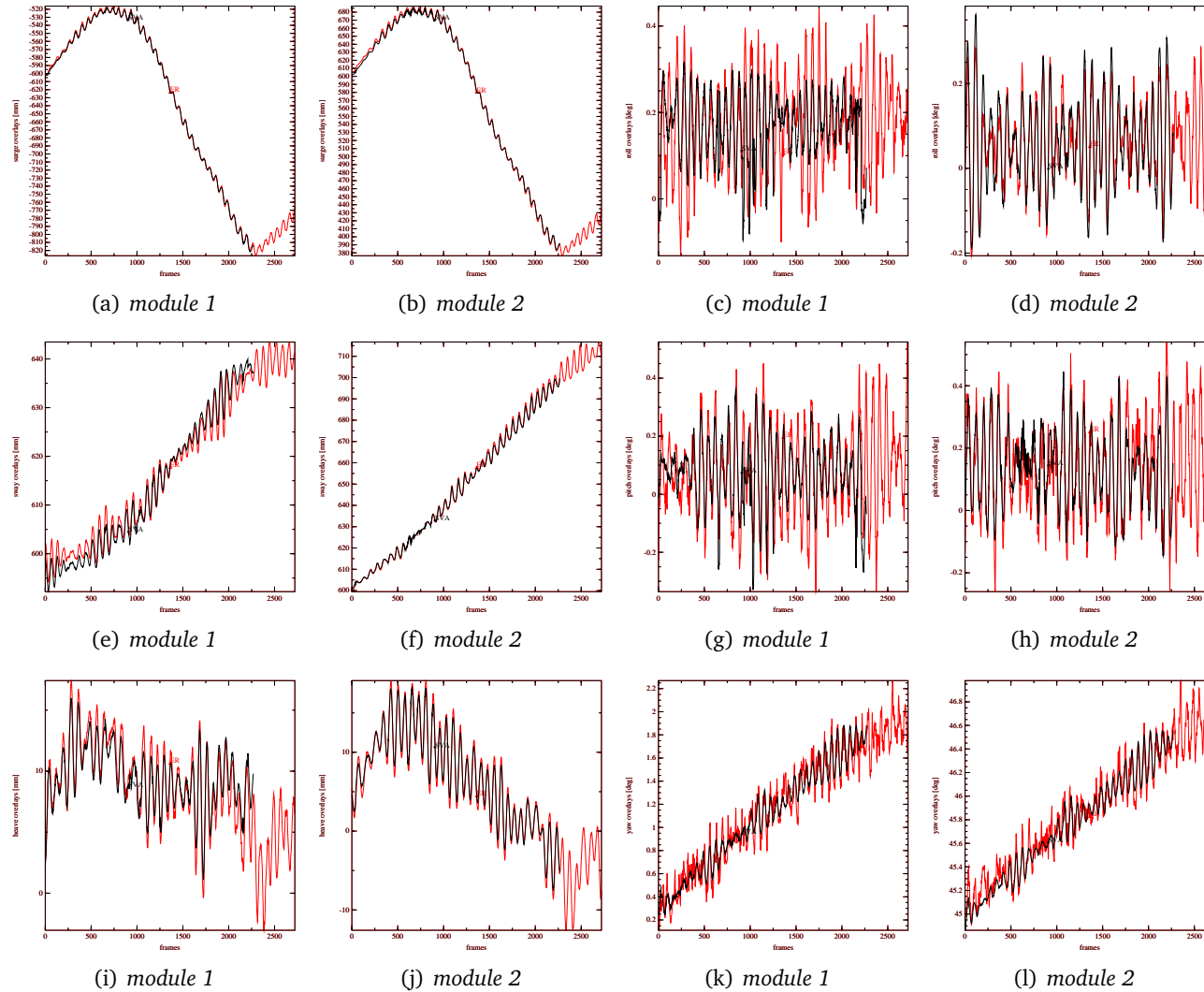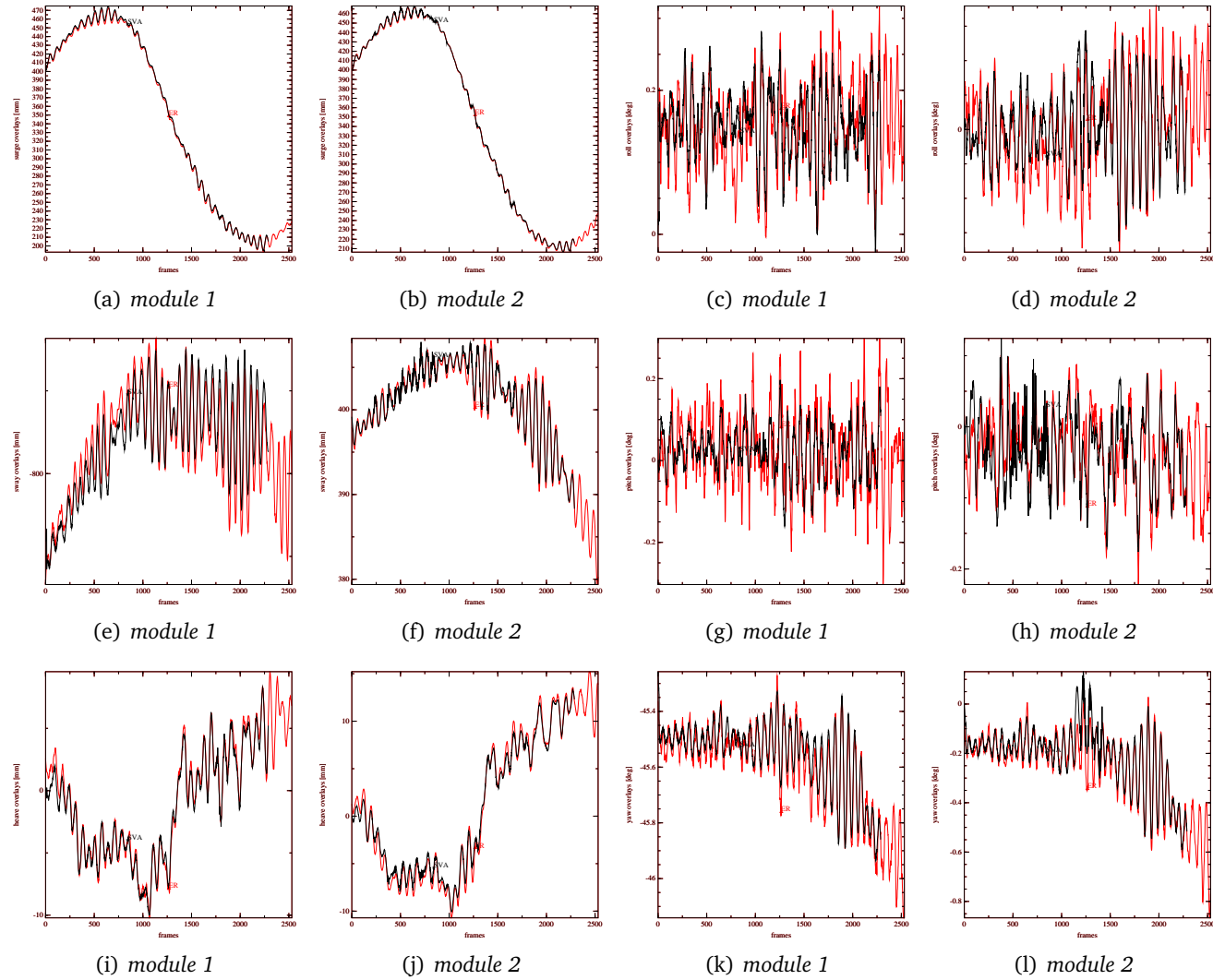
Figure 4.13: Measurement results (graphical presentation) of *Series 3* in terms of ship motions. The signal response in red comes from the developed measurement system, whereas the black refers to MCS. Surge in (a), (b), sway in (e), (f), heave in (i), (j), roll in (c), (d), pitch in (g), (h), and yaw in (k), (l). *Module 1* and 2 are defined in Figure 4.8.

Dynamic case

In the dynamic case, a single measurement was evaluated. The MCS data in the irregular wave condition, that is when the camera systems moved as the measurement was taking place, did not provide the motion corrections thus it was not possible to confront the two results. As a workaround, we picked eight distances on the platform grid (see Figure 4.14), and as the imaging system was performing the dynamic referencing, we investigated whether their lengths comply with their expected values.

Throughout the measurement, a change of the cameras position was registered seventy times. As the motion was not continuous, the shifts in position between the frames in which the change took place were interpolated with a second order polynomial (see Figure 4.15). The validation distances were checked solely at instances of the change occurring (see Figures 4.16 and 4.17)
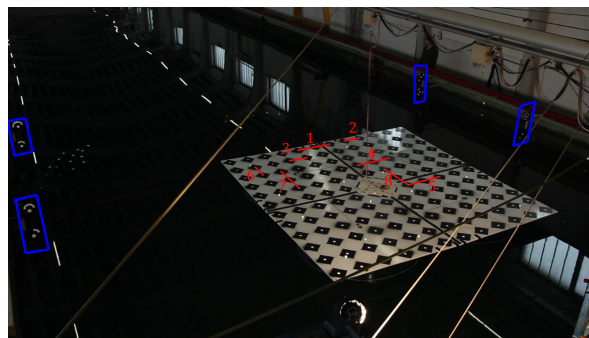
Figure 4.14: Dynamic referencing. The motion of the camera system was retrieved thanks to the static points marked with blue. The enumerated and marked with read are the distances that served to evaluate the method.

(a) *motion along X-axis*   (b) *motion along Y-axis*   (c) *motion along Z-axis*
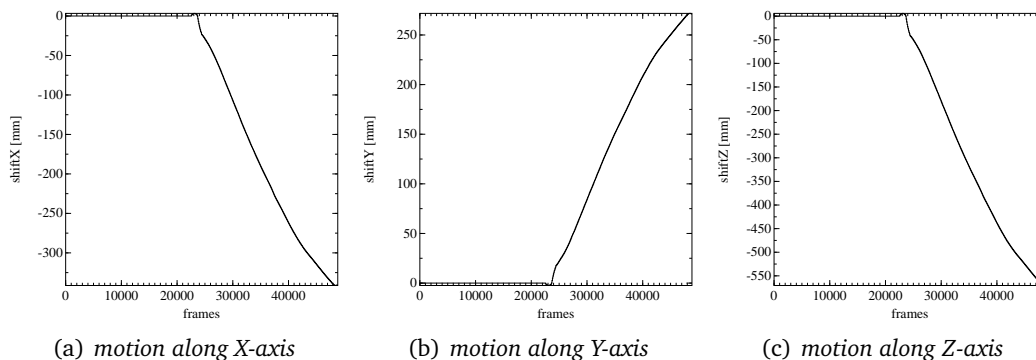
Figure 4.15: Motion of the camera system computed by dynamic referencing. The coordinate system was defined in the perspective center of one of the cameras, in its initial position.

Scalebar	$distance$ [mm]	$RMSE$ [mm]
1	674.0	0.5
2	334.4	0.3
3	333.4	0.8
4	676.3	0.8
5	678.8	0.4
6	341.9	0.8
7	680.2	1.8
8	673.1	0.4

Figure 4.14

Figure 4.16: The dynamic measurement presented in figures. The root mean square error (RMSE) was calculated confronting the distances prior to the motion occurring, and computed in the dynamic bundle adjustment.
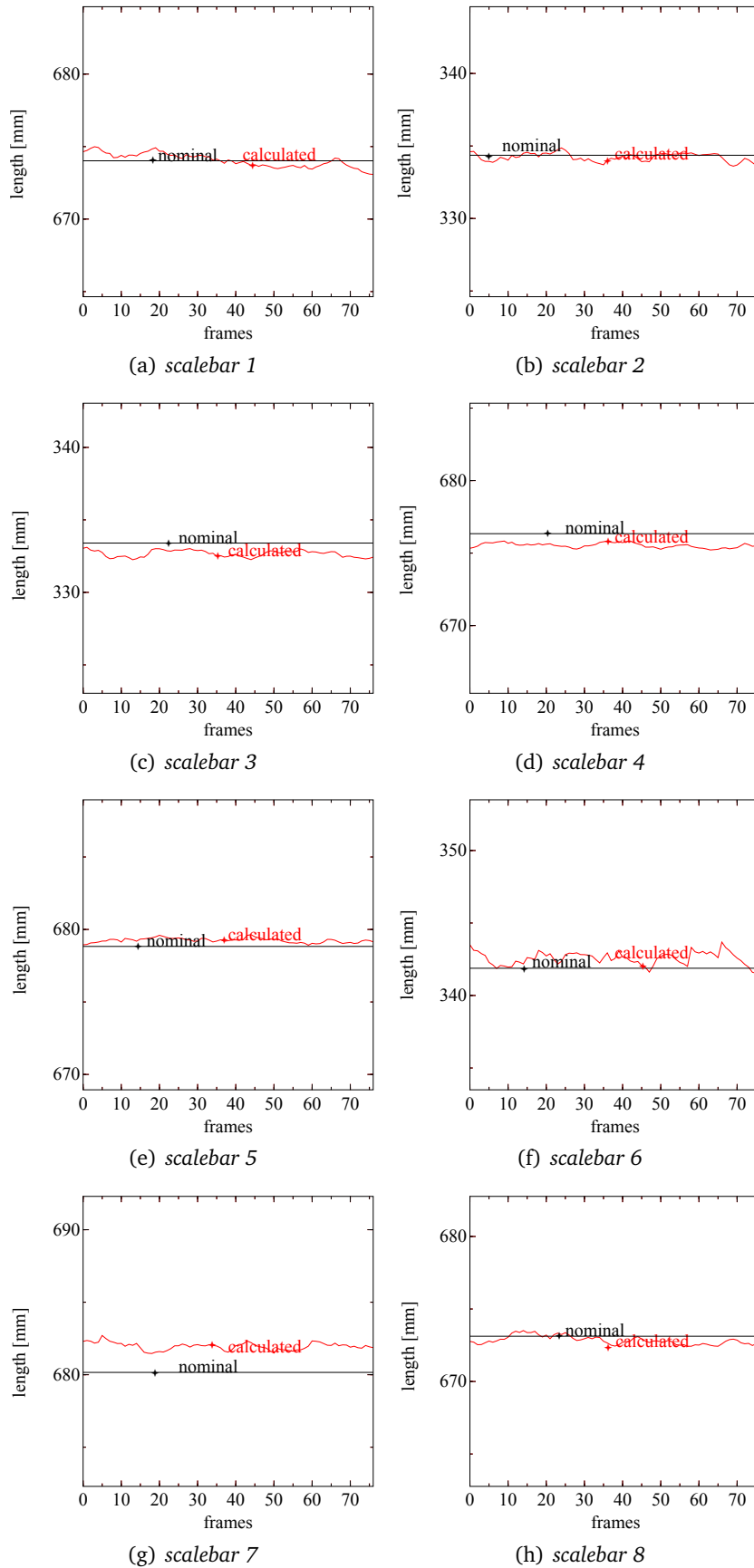
Figure 4.17: Graphical representation of the result in the dynamic camera case. The red signal correponds to the scalebar lengths in subsequent frames, calculated by the the developed measurement system, whereas the black is its nominal length.

4.5 Discussion

A three-camera imaging system was proposed. The adopted cameras were of low-cost make (total price < 5k EUR), and intended for consumer use. The motivation behind selecting equipment of this sort to carry out measurement activities was to quantitatively assess its feasibility as a photogrammetric tool, or to stress the time component — as a videogrametric tool.

Camera related issues   The camera specific problems were the side-effects of the camera primary functions given by the manufacturer. The major predicament concerned the fact that the Software Development Kit (SDK) included no established functions to start and stop a video acquisition, and consequently no established methods to synchronise a multi-camera acquisition. External, commercial solutions [*Breeze Systems*] were required to overcome the limitation. Same holds true for the synced operation of the illumination equipment with the cameras, leaving the user merely with the continuous illumination as a choice. Because a custom made solution is a puzzle one must assemble, it may be discouraging for a potential user with no background in the field. If it is the case, a turnkey motion capture system may be preferred. Lastly, one shall consider the fact that a consumer grade equipment allows to collect a maximum of 30 frames per second at Full-HD resolution, thus resolve a motion of a frequency less than 15Hz. An MCS, in general, offers a wider range of acquisition rates.

In view of the author, the apparent loss of accuracy due to unstable mechanical connection between the sensor and the lens, plays a minor role. The self-calibration is anyway a routine before the measurement taking, plus during the acquisition the cameras remain untouched. Even in the dynamically referenced bundle adjustment, there is little probability of the IO parameters and lens distortions changing.

Camera non-related issues   The camera independent obstacle stemmed from the workplace environment, with limited freedom and options of control, i.e. occlusions, constrained space and access, presence of objects generating abundant specularities. The hardships consequently influenced the survey outcome for the network design could not be planned assuring highest precision, whereas the glitter and wandering specularities precluded the point detection, centroiding and tracking. Yet, albeit the restrictions equally applied to both measurement equipments, the custom solution proved superior in one aspect. Unlike the MCS which as the output delivers exclusively 3D coordinates, our system stored the videos. The tracking issue could be handled in post-processing, by applying the right algorithms (see Kalman predictor in Section 4.3.3). The MCS presumably lacked measures to counteract the timely occlusions given the blunders in the data and partial registration of points.

Accuracy   The points were triangulated in a bundle adjustment with the camera perspective centers regarded constant. The resulting precision measures ($\sigma_{x,y,z}$ in Figure 4.9) may render too optimistic because freed from the uncertainty of the IOR and

EOR parameters. Albeit with strong imaging networks this difference remains unnoticeable (see the *Limiting Error Propagation* in Fraser [1987]), more attention was paid to the aspect of accuracy as the evaluation measure, by confronting the results with objective data, i.e. independent and of higher accuracy.

The chief factor determining the ultimate precision/accuracy, which should be thoroughly planned in advance, is the network design. The corollary effects of not following the design rules are well observed in Figure 4.9. In all surveys the precision/accuracy measures on X- and Z-coordinate came out worst. The explanation is in the distribution of the cameras within the measurement space and with respect to the surveyed platform. A large base yielding good intersection angles is guaranteed solely along the Y-coordinate. Along the X- and Z- there is practically no parallax. Note in Figure 4.4 that the Y-coordinate runs across the model basin, the Z runs down, and the X towards the platform.

Further repercussions of the suboptimal imaging network are apparent when confronting the $\sigma_{x,y,z}$ of the two tested modules in Figure 4.9. The $Module2$ was positioned closer to the bridge carrying the cameras, hence captured at a greater scale with respect to $Module1$. The $Module1$ precision values are equivalently rendered more optimistic.

Looking at the graphical superposition of the validated and true signals in Figures 4.11, 4.12, 4.13, no systematism can be registered, suggesting that the figures of accuracy contain only the random errors. The translation part of the *ship motions* is clearly better adhered to the true signal, than the rotational components. It is justified by the fact that the shifts were computed comparing the centroid of the platform, that is an averaged value. *Yaw, pitch, roll* were derived from non-redundant data, thereby, the random errors are more pronounced. Besides this, the increased noise levels are certainly due to the artefacts caused by the lossy H.264 compression.

The dynamic referencing did not influence the quality of the produced results. The controlled distances were of lengths ranging between 300 and 700mm, arranged along the X and Y axes. The $RMSE$ in Figure 4.16 did not indicate the dependencies on the location or length of the distance. The order of discrepancies was within the anticipated measurement precision.

Water surface shape reconstruction

This chapter finalizes the experimental part of the dissertation. To start with, techniques for water surface measurement are described. Section 5.1 reports primarily on methods used in industrial implementations thus does not exhaust all existing water measurement approaches *per se*. Over the course of the last few decades, the computer vision community has been an important contributor to much development in analysing the water shape. Nonetheless, the conceived theories rarely go beyond the proofs of concept. For an overview of the principal approaches, the reader is referred to Subsection 2.4.4.

In Section 5.3, the thesis's contributions are disclosed. The two developed methods regard the water as either diffuse or specularly reflecting surface.

5.1 Related work

Historical review

Attempts to characterize the water surface with optical methods date back to the beginning of the 20th century [Kohlschütter 1906; Laas 1906]. The interest in a quantitative description of the surface with light came from the field of oceanography and the use of photography to map the coastlines. This prompted further applications, namely the use of photography to quantify ocean waves, and exploit these parameters in e.g. ship-building to engineer structures of the appropriate strength. Photo-theodolites carried on-board a ship were the first instruments to be used for this purpose. Over the course of years, they underwent an evolution from being capable of capturing a single stereo pair to observing time sequences, nonetheless, the manual data interpretation remained the bottleneck and the method was regarded impractical [Kiefhaber 2014; Waas 1988].

There was a strong evidence that photography could be useful in ship-building industry, but the advancement of technology was impeding its use on a large scale. The data evaluation was too costly and labour intensive to perform, as well as the understanding of the seakeeping abilities still not satisfactory. Despite the Froude's breakthrough in

scaling laws[1] in the mid 19th century, and the subsequent erection of the world's first towing tank, a hundred years needed to pass for the relationship between the ship's behaviour in laboratory and real environments to be conceived [Journee and Pinkster 2002]. The mid 20th century advances in naval architecture, and the rising power of the computer led to the development of methods of numerical simulations. In particular the 1970's were the turning point thanks to offshore oil and gas production. Expected wind- and wave-induced ship motions and loads could then be efficiently calculated by a computer program. Yet, as the computer alone depended on the theoretical models behind it, model tests were necessary as before, especially when new concepts were tested [Faltinsen 1993]. For photogrammetry, the digital era, invention of charged couple devices (CCD) and the computers with large memory capacity meant that this technology was no longer merely spoken of in the books, but could finally be deployed to observe time-dependent phenomena. Manual work gave way to automated processing, and the human operator interaction could be fully or partially eliminated.

In parallel, the same drivers disseminated optical methods among other applications — in river engineering and the oceanographic domain. Understanding the river flow allows for a better riverbed management and mitigation of floods through combined fluid dynamics modelling and experimental testing. Additionally, the knowledge of the dispersive processes gives an insight into the way pollution and sediments are transported. The latter should be regularly monitored to assure good navigation conditions for ships [Chandler et al. 2008; Godding 2003; Mulsow et al. 2008; Stojic et al. 1998].

In the coastal zones, optical methods became a good alternative to in-situ measurements which require substantial logistical commitments and offer low spatial as well as temporal resolution. The dynamics of the water, hence the energy it carries, influence the nearshore morphology which is of significance for both the coastal communities and the marine infrastructure, e.g. wharfs, mooring systems [Aarninkhof et al. 2003; Adams and Pos 1984; Redondo et al. 1994; Santel 2006].

Last but not least, the roughness of the ocean's surface is meaningful from the viewpoint of exchange processes with the atmosphere. The oceans, accounting for two-thirds of the Earth surface, contain high concentrations of carbon dioxide and thereby influence the global climate system through the evaporation and absorption phenomena. The optically measured wave slope can be used to parametrize the relationship between the surface roughness and gas transfers, as well as the wind speed and direction [Jähne et al. 1994; Kiefhaber 2010].

Volumetric observations of water with e.g. particle image velocimetry (PIV) are not reviewed hereafter.

Optical methods

One can distinguish between three groups of measurement methods, each exploiting different characteristics of the water medium. The water surface can be regarded (a)

---

[1]Froude's scaling laws allowed prediction of the real-size ship's resistance based on the performance of a scaled model.

diffuse-like (quasi lambertian), (b) mirror-like (specular reflection), or else (c) as a transparent medium (refraction). When the wavelength of the incident radiation is smaller or equal to that of the surface roughness, the surface will appear as diffuse. On the other hand, smooth surfaces will produce specular reflections and their appearance will be highly dependent on the position of the observer as well as the radiation source [Nayar et al. 1991].

Water as a diffuse surface. In practice, there are two conditions when the water is rendered diffuse-like. First, when the surface is disturbed with small waves (e.g. rough sea in the coastal zone), second, when artificial targeting is employed. The most common targeting techniques use physical material like powder, styropor, oil, or optical projections in forms of a laser sheet, grid of points, or sinusoidal patterns [Adams and Pos 1984; Cobelli et al. 2009b; Maresca and Seibel 1976; Mulsow et al. 2008; Piepmeier and Waters 2004]. Specular reflections are inevitable and are often the source of errors in the estimated depths. To avoid the corrupted measurements, specular highlights can be: (a) removed in the image preprocessing step, (b) eliminated in multiview setups during the processing[2] [Bhat and Nayar 1995], (c) filtered with the help of either polarized or chromatic filters[3] [Wells and Danehy 2005], and (d) in industrial photogrammetry of rigid objects, attenuated through the use of special targets e.g. doped with fluorescing dye, that respond to a wavelength other than the wavelength of the specular reflections [Black et al. 2003].

Alternatively, the water being itself a source of infrared radiation can be observed with thermal cameras. Because the heat distribution is heterogeneous across the surface, it provides a good base for the correspondence search (add ref). Surface treated in this way can be measured with classical stereo- or multiview photogrammetric approaches (Section 2.3). If qualitative results are expected, it is sufficient to acquire single images and proceed with data evaluation in the image space only [Lippmann and Holman 1992].

Water as a specular surface. Sometimes it is advantageous to exploit the inherent optical characteristics of water, i.e. the total reflection and refraction, for measurement purposes. The water as it is, rarely gives enough pattern to unambiguously match homologous points in sets of images. Besides this, contaminating the liquid with physical material is cumbersome because it becomes (a) unfeasible for large areas and field survey, (b) difficult to keep a homogeneous point distribution, and (c) it may influence the response of the water by interacting with it. In such situations, and depending on the working environment — whether in a lab or out in the field — it is possible to derive the surface shape by mere observation of a reflection of a source light or a refracted pattern whose distortion corresponds to the surface slope and height.

---

[2]Appearance of glints in images is view dependent. With 3rd or n^{th} view, every identified feature can be verified and glints can be eliminated. The method is apt for scenes with single or few glints.

[3]The filters are mounted in front of the camera lens hence there is no restriction on the number of present glints.

Cox and Munk [1954] pioneered the characterization of water surface slopes with a reflection based method. Their motive was to analyse slope distributions under different wind-speeds through observing the Sun's glitter on the sea surface from an aerial platform. The slopes' frequency distribution was derived from the brightness of the glitter pattern in the images and the method was therefore said to be susceptible to systematic errors stemming from multiple reflections, shadowing effects, non-uniformly bright sky or white caps caused by high speed wind. Variations of the Cox and Munk method include: replacing the natural illumination with one or more artificial light sources, a.k.a. Reflective (Stereo) Slope Gauge or RSSG [Schooley 1954; Waas 1988] using entire clear or overcast sky to derive surface slope information for every point in the image, a.k.a. Stilwell photography [Stilwell 1969].

Jähne et al. [1994] revisited these techniques to study their actual suitability for inspecting the specular ocean. Deriving from the shape-from- ... paradigm, the authors point out the practical limitations of the shape from reflection[4] method. First of all, because the 2D slope information is mapped to a scalar-valued pixel intensity, calculation of the surface orientation is ambiguous. Secondly, when executed with an extended light source, the main obstacle is the narrow range of the observable slopes, or otherwise the size of the wide-angle illumination needed to extend its scope becoming infeasible to put into practice. Exploiting the daylight luminance renders a viable solution to that problem, however, it imposes the rigid demand of clear sky hence precluding the survey outside rare weather windows. All in all, such shape from reflection approaches are not recommended for a systematic study of waves.

However, a combination of stereo- and reflection-based techniques (RSSG) was proven to be a sound way to characterize not only the surface slopes but also the water heights. In a typical stereo setting, one is faced with a bias in the corresponding features seen by the left and the right cameras. The reason being that the cameras will record a change from the spots on the water whose normals' are in the line of sight of the given cameras. Naturally, the steeper the observed wave, the less the systematic error. Using the Helmholtz reciprocity principle, i.e. upgrading the method to employ light sources placed right next to the cameras, eliminates the correspondence ambiguity. The identified features in image spaces are then bound to be a unique feature in the object space [Kiefhaber 2014; Waas 1988].

In the field of computer vision the set of solutions arising from calculating the surface slopes from a single image is narrowed down by enforcing a constraint, for instance a surface curvature constraint or assumption on the surface material. A more straightforward solution would be to disambiguate the incorrect solutions by engaging several "wandering" illumination sources. In the case of diffuse-like materials, the method is known under the name of photometric stereo[5], first exposed by Woodham [1980]. Three years after, Ikeuchi [1981], driven by the industry and the need to map the shape of glossy objects, successfully adapted the method to tackle specular surfaces. Schulz [1994] further modified the method to work in natural lighting conditions, thereby mak-

---

[4]Shape from reflection is an adaptation of the shape-from-shading method on specular surfaces.

[5]Photometric stereo is sometimes referred to as multiple image shape-from-shading.

ing it truly apt for real world scenarios. He replaced numerous illumination sources with numerous cameras and a range finder. The approach consists of a surface orientation step and a surface propagation procedure, which converge to the correct surface by minimizing the difference between observed and synthesized image irradiance values. The shortcomings of the algorithm are its many inputs as well as the fact that the authors illustrate only simulated results.

A myriad of computer vision approaches have subsequently come to light. They commonly divide between shape from specularity and shape from distortion approaches, and build upon the aforementioned principles, e.g. inspect single or multiple highlights with static or moving observer and emitter; observe known or unknown intensity patterns reflected from mirror-like surfaces; directly measure the incident rays; exploit light polarization; make assumptions on the surface's bidirectional reflectance distribution for curvature retrieval. Their superiority certainly results from the enhanced computing power, the weakness is the study cases that work on glossy and specular rigid objects but never the water. For a good overview of the techniques refer to Ihrke et al. [2008]; Morris [2011].

Water as a refractive medium. Refraction-based techniques are more complex due to the fact that the light path is dependent not only on the surface normal, but also the medium's refraction index. The development of shape from refraction goes hand in hand with the shape from reflection methods and therefore has an equally long history. Cox [1958], again, first experimented with light refraction to derive surface slopes from an intensity gradient emerging from beneath the water surface. Today, a successor of this technique — called imaging slope gauge alternatively — shape-from-refractive-irradiance, is considered a highly reliable instrument for measuring wind induced waves in laboratory environments. The measurement setup comprises of a submerged light source generating the intensity pattern and an image sensor placed outside the water to record the outcoming rays. The light source locates one focal length behind a Fresnel lens and thereby directs the rays of one inclination onto the same spot of the illumination gradient. The intensity gradient may be 1D or 2D, coding the positions in one or two dimensions. Given an idealized camera i.e. at infinite distance and pointing vertically towards the water surface, an intensity value observed at the image plane is distance-independent[6] and allows for the computation the 2D wave slope. In contrast to the reflection-based techniques, refraction of a pattern through the water maintains a quasi linear relationship between the slope and the image irradiance [Jähne and Riemer 1990; Jähne et al. 1994; Rocholz 2008; Zhang and Cox 1994].

White light can also be replaced with an active laser lighting. The laser offers a greater temporal resolution at the expense of a lower spatial resolution. First examples of laser imaging gauges employed a single beam, thereby delivering information about a single slope [Hughes et al. 1977; Sturtevant 1966]. Over time they evolved to scanning systems that could capture spatial information over an area [Bock and Hara 1995].

---

[6]This condition is achieved by employing a telecentric lens. The image to object projection is then taken as an orthographic one.

Laser slope gauge is used in the laboratory and field environment, however, most apt for characterization of longer waves as the literature reports.

In the field of computer vision, dynamic transparent surfaces have been analysed with single and multiimage approaches. Murase [1992] first introduced a single view shape from the motion of a refractive body. He proposed a method for reconstructing the shape of an undulating water surface by inverse ray-tracing. The underwater pattern was a blank sheet with randomly scattered black dots, whose apparent motion was tracked with optical flow methods. By assuming that the water's average slope is equal to zero i.e. it regularly oscillates around a flat plane, the author infers that the center point of a point's trajectory in time shall lie on that average shape. This evidence permits the computation of surface gradients for the remaining vertices of the trajectory, which, when later integrated, establish the final reconstructed shape. The algorithm is fully functional with no external knowledge, nevertheless when the medium's refractive index and the pattern to water surface height are given, results of real scale are obtained.

On the other hand, Morris and Kutulakos [2011] developed a method that recovers complete information on 3D position and orientation of dynamic, transparent surfaces using stereo observations. The authors formulate an optimization procedure that minimizes a so-called *refractive disparity* (RD). In short, for every pixel in an image that refracts from an underwater pattern, the algorithm finds a 3D point on the water surface that minimizes the conceived RD measure. This measure expresses the Snell's law of refraction by enforcing that the normal vector of the computed 3D position must be equal when computed from the left and the right images of the stereo.

As a conclusion, the two flag examples of computer vision approaches are limited to use in laboratory conditions due to (a) the need to locate a reference pattern under the water, and (b) the demand of clear water for reliable a pattern to image correspondence retrieval. For more information refer to Ihrke et al. [2008]; Morris [2011].

5.2 Problem definition

Following the material of Chapter 4, the algorithms presented below deal with the cause of the platform motion, that is the water waves. The task was to reconstruct the instantaneous shape of the water surface in the vicinity of the floating platform. It was accomplished in a twofold manner, (i) by observing few physical targets floating on the top of the surface (*Method 1* in Subsection 5.3.2), and (ii) by observing the apparent motion of specular reflections coming from points of known coordinates (*Method 2* in Subsection 5.3.3). In either case, the transfer from point-based measurements to a surface was possible thanks to the prior knowledge on the wave excitation.

The waves were generated with a linear mechanical arm, occupying the entire width of the basin. The wave propagation followed a single direction along the basin's longer axis. As such, the arm moved at equal time intervals producing sinsoidal waves at various but constant frequencies (f), periods (ω) and wavelengths (λ), see Figure 5.1.

The employed imaging system included the same three low-cost cameras as describes in Subsection 4.3.1. All the algorithm routines were implemented in own software tools.
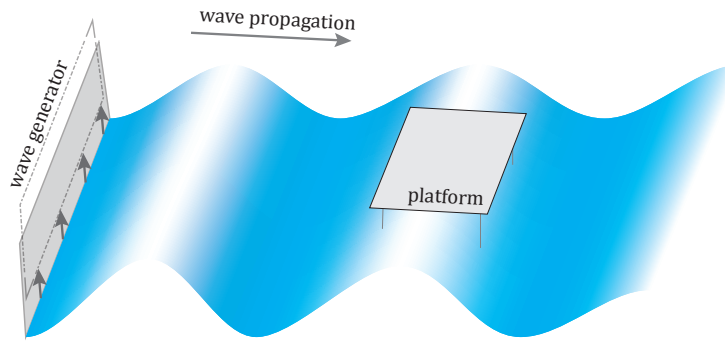
Figure 5.1: A simplistic view of the model basin. The platform is placed at a distance from the wave generator (on the left) and parallel to the wave propagation direction.

5.3 Adopted methods

5.3.1 Preliminaries

The travelling 3D sine wave[7] is deemed

$$\mathbf{y}(t_i) = \mathbf{y}_i \tag{5.1}$$

$$\mathbf{z}(t_i) = A \, \sin[\, 2\pi \, (\frac{t_i}{T} - \frac{\mathbf{y}_i}{\lambda}) + \phi\,], \quad i = 1 \ldots N \tag{5.2}$$

where ϕ is the phase of the wave-front, and N is the duration of the measurement in seconds. The travelling sine wave is a *special surface* and as such it is advantageously modelled in a *local coordinate system* (LCS), that is parallel to the wave propagation direction, and shifted to the mean level of oscillations (see Figure 5.9). In order to link the LCS with the *global coordinate system* (GCS) where the cameras are defined, and a 3D spatial similarity transformation (the scale factor is unity thus the transformation reduces to a 3D rigid transformation; see Equation (2.8)) is formulated. The parameters of this transformation (i.e. three components of the translation vector and three rotation angles) can likewise be included within the adjustment.

The methods invented devise the above defined model in a least square approach, that is trying to minimise the residuals between the nominal observations, and the observations predicted by that model. To solve the non-linear least squares problem, one must know the starting values of all parameters involved. Presented below is the strategy for retrieving the wave amplitude A, the wavelength λ, period T, phase-shift ϕ and the rigid transformation of the wave LCS to the camera GCS. Experience has shown that unless the amplitude is infinitesimally small, and the wave infinitely long, even very rough estimates of the unknown wave parameters assure convergence of the system of equations.

---

[7]Note that the shift term H_0 in the Equation (5.2) is omitted. This is possible if the translation from the $GCS \to LCS$ already compensates for that shift.

Derivation of approximate wave parameters

The wave amplitude can be (i) estimated visually, on-site, during the measurement taking place, (ii) recovered from the wave probes, as these are a common place in any ship model basin, or (iii) derived from image-based measurement, provided there are observed real points found on the water surface. When the image-based approach is undertaken, the triangulation step must follow to obtain the 3D coordinates of the real points (see Subsection 5.3.2). The amplitude can be recovered with the *complex amplitude demodulation* method (AMD), or merely by removing the trend from a point's response and taking the halved maximum bound to be the starting A value. The collateral benefit of the AMD is that, were there a varying amplitude signal, a slope instead of a horizontal line outcome would be observable. Accordingly, the A shall be replaced with a (linear) time varying function, the parameters of which may be engaged in the total adjustment too.

Similarly, the value of the period T can be approximated either from on-site visual impressions, using the image data in post-processing, or taken directly from the wave probe data. The dominant period is then restored with the help of the spectral analyses i.e. the periodogram power spectral density estimate.

The wavelength might be inspected visually, or with the help of the method presented in Section 5.3.2. The wave probe devices, being single point based measurements do not deliver enough data to indicate the length of the travelling wave.

The remaining wave parameter is the phase-shift. Its computation requires a real point floating on the water surface, or otherwise, the wave probe data can become useful if its position is known in the reference coordinate system. If one moves the CS to that point, the term $\frac{y_j}{\lambda}$ in Equation (5.2) cancels out. If one further takes the starting frame $t_1 = 0$, also the term $\frac{t_1}{T}$ is gone, and the phase-shift can be computed from $\phi = \arcsin \frac{z_j}{A}$, where j denotes the point at the origin of the translated CS. A slightly more elegant way to solve the equation for the initial phase-shift is to use, again, the demodulation technique.

As for the wave transformation parameters $(\omega, \Phi, \kappa; \mathbf{T})^{GCS \to LCS}$, one usually tries to define the global system that is *quasi* parallel to the wave propagation direction, or in best case aligns with it. If so, the rigid transformation parameters can be set to zero values at the beginning of the adjustment, and receive their corrections, which compensate for the inaccurate alignment. Otherwise, the image data and the scene context can be exploited to find the transformation relating the two coordinate system by identifying e.g. the minimum of three common points or a 3D line and a point.

Optimisation technique

In the computational part, the Gauss-Markov least squares adjustment with conditions and constraints was adopted. The adjustment workflow proceeds in repetitive cycles of five steps (see Figures 5.6 and 5.10), i.e. (i) generation of current approximate values of all unknown parameters, (ii) calculation of the reduced observation vector,

(iii) calculation of the partial derivatives of the functional model with respect to all current parameters, (iv) construction of the normal equations and (v) solution of the system of equations.

The partial derivatives that make up the Jacobian matrix are always evaluated at current values of parameters. The derivatives of selected constraint equations can be found in Appendix A.1, for the condition equations the reader is referred to the literature, e.g. Kraus [1997]; Kraus [2007]; Mikhail et al. [2001]. If the system of equations includes both condition and constraint equations, it does not fulfil the positive-definite requirement put by e.g. the Cholesky decomposition. Indeed, on the diagonal of the equation matrix in *Method 2*, zero entries are present. The vector of solutions is then retrieved with the Gauss elimination algorithm.

Assessment of the precision of the output parameters is contained within the covariance matrix, i.e. the inverse of the normal equations. For large systems, one tries to avoid performing that operation due to the excessive memory-load and computational time entailed. In our implementation, quick recovery of the diagonal entries of the inverse matrix were obtained thanks to the *principal pivot transform* (PPT) and its properties providing the least squares solution [Tsatsomeros 2000].

The underlying model equations appear in the form of the Linear Substitute model (see Section 3.1) which involves finding the parameter estimates iteration-wise. The stopping criterion was defined as $\mathbf{l}^T\mathbf{l} - \mathbf{v}^T\mathbf{v} < \epsilon$, with the ϵ set to 0.0001 [Kraus 1997].

5.3.2 Method 1 — diffuse surface

The following method sees the water as a diffuse surface. It is converted to such, owing to artificial targeting. A set of retro-reflective targets floated on the water surface, and were tracked during the measurements (see Figure 5.2). Reconstruction of a singular point provides for all but one wave parameter — the wavelength λ. Combining the responses of minimum of two such points allows for the possibility of recovering also the remaining λ. It is presumed that the 3D data has already been transformed to the *LCS*, and to shorten the discussion it is excluded from the mathematical model. A complete description on how to include this information in the model is given in Subsection 5.3.3.

Since the parameters are found in a least squares approach, the discussion commences with the initial parameter retrieval. In the next step, the adjustment mathematical model is outlined (see also Figure 5.6). Evaluation of the result continues in Section 5.4.

Image-based approximate wave retrieval

The starting point is to measure the targets in images, for instance with the methods described in Subsection 4.3.3. The 2D image measurements are then transferred to 3D space in a regular bundle adjustment, and exploited to recover the initial wave parameters. The developed pipeline is fully automatic and summarises in the following order
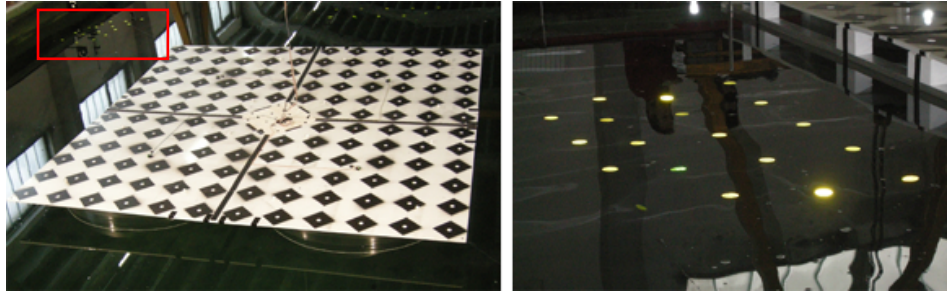
- clustering of 3D points,

Figure 5.2: Left: physical, retro-reflective tartgets floating in vicinity of the platform. Right: a close-up of the targets. The targets are interconnected with a string so as to avoid their collision and the dispersion.

- coupling of neighbouring *clusters*,

- calculation of mean A, T, and ϕ from the *clusters*,

- calculation of mean λ from the couples of *clusters*.

The reconstructed targets in time are considered an unorganised point cloud thus their clustering is carried out up front (see Figure 5.3). The retrieved clusters are equivalent to the response of a single target floating on the water surface. The clustering *per se* is not required as this piece of information is already carried within the point naming convention. Nonetheless, because this may not always be the case, the clustering is a default operation. It creates boundaries in 3D space based on some measure of proximity, in our case the Euclidean measure was chosen. The algorithm was invented by Rusu [2009], and implemented in the *Point Cloud Library (PCL)*.

Now, the coupling establishes a neighbourhood relationship between closest clusters (see Figure 5.3). Given a *kd-tree* representation of all clusters' centroids, the algorithm searches for the neighbours within a desired radius, and ascertains that the selected pair has an offset along the wave propagation direction. This later permits for the computation of the wavelength λ. The selected radius should not be too small, but cannot be greater than the wavelength to be able to resolve its length.

The mean A, T, and ϕ are found within each single cluster. The amplitude A is calculated as the mean halved distance between the maximum and minimum heights. The period T is calculated from the power spectrum estimation, also known as the *periodogram*. The periodogram takes a signal of length N, that is a cluster, computes its discrete Fourier transform (FFT) and estimates the power spectrum at $N/2 + 1$ frequencies. The principal frequency present in the data is the one manifesting the largest power spectrum density.

To find the mean λ, the requirement is to know (i) the period T, (ii) the direction of the wave propagation, and (iii) the relation between the GCS and the *LCS*, where the wave is defined. For every couple of clusters, one counts how much time Δt a wave crest

takes to travel between the clusters (see Figure 5.4). The distance Δd (in *LCS*) between them is known, and so is the period T, hence the wavelength estimate results from the trivial proportion

$$\frac{\Delta d}{\Delta t} = \frac{\lambda}{T}.$$ (5.3)
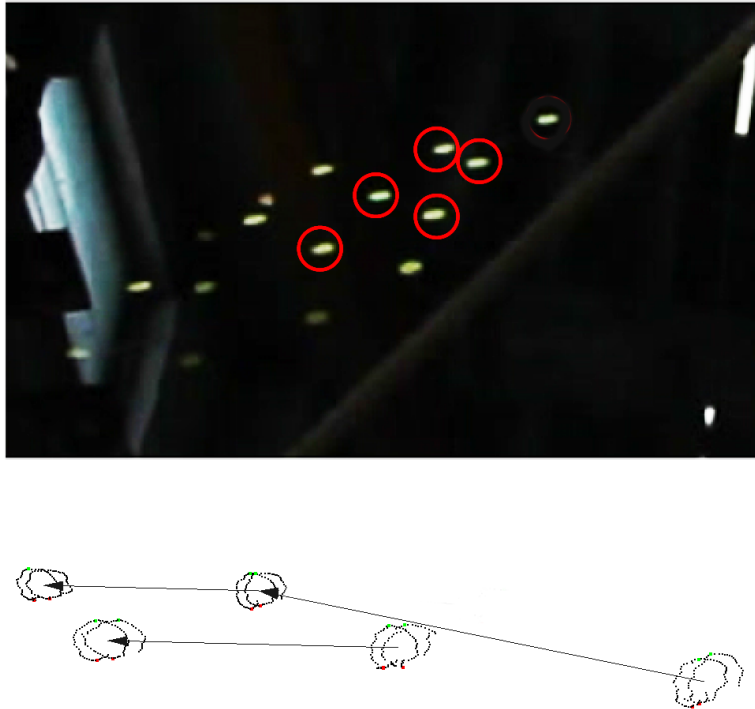
Figure 5.3: Image-based approximate wave retrieval. Top: physical targets visible in 2D images. Bottom: perspective view of the five recognised *clusters* in 3D. The arrows indicate established cluster pairs.

The mathematical model

The functional model describes analytically the prior knowledge on the wave shape expressed in Equation 5.2. The modelling part, unlike in the *Method 2* in Subsection 5.3.3, is formulated purely in 3D space. The point triangulation is treated as a separate and unrelated phase, even though the image measurements indirectly contribute to the eventual outcome, and a joint treatment might be suggested. As a matter of fact, the 3D space is reduced to a 2D space, building on the fact that the transformation from *GCS* to *LCS*
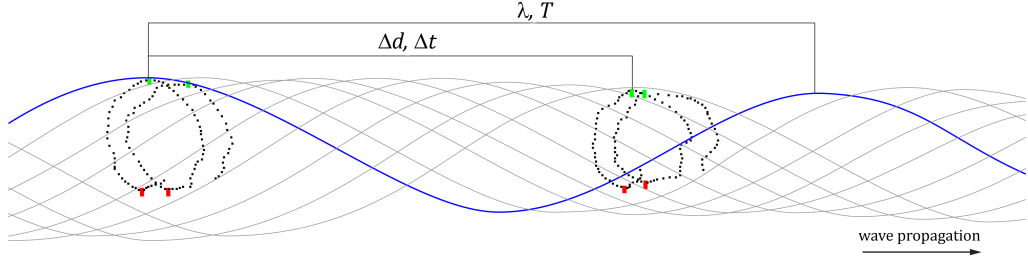
Figure 5.4: Image-based approximate wave retrieval. For this *cluster* pair, the wave crest takes eight frames (Δt) to travel between the neighbours. If the distance between the clusters is $\Delta y = 10cm$, and the period $T = 12$, then the wavelength $\lambda = 15cm$.

has taken place in advance, and the x-coordinate can take arbitrary values. All parameters are present in the adjustment as observations and unknowns; see the functional model parameters in Figure 5.5, and the adjustment workflow in Figure 5.6.

| Model	Parameters
Wave	A, λ, T, ϕ, $\left(\mathbf{T}^{GCS \to LCS}, \omega, \phi, \kappa^{GCS \to LCS}\right)^*$, yz^{real}

Figure 5.5: The functional model of the least squares adjustment in *Method 1*. The wave parameters correspond to the shape defined in LCS — local coordinate system. $\mathbf{T}^{GCS \to LCS}$, $\omega, \phi, \kappa^{GCS \to LCS}$ are the parameters of a rigid transformation that bring the wave shape to GCS - global coordinate system. *denotes that in the developed implementation parameters were treated as constant.

The stochastic model is formalized within the weight matrix and conveys the observation uncertainty. The matrix holds non-zero elements on its diagonal, and they may take the following form

$$w_i = \frac{\sigma_0^2}{\sigma_i^2}. \tag{5.4}$$

where σ_i signifies the *a priori* standard deviation of an observation, and σ_0 is the *a priori* standard deviation of a unit weight. Within the experiments, the σ_0 was set to unity, whereas $\sigma_A = 2mm$, $\sigma_T = 0.05s$, $\sigma_\lambda = 100mm$, $\sigma_\phi = 0.25rad$. These values are rough and rather pessimistic estimates of the uncertainty of the approximate parameters.

Figure 5.6: The least squares adjustment workflow of *Method 1*. The cycle repeats until the stop criterion is reached. $\partial \mathbf{x}$ is to denote the entire vector of parameters, while the exclamation mark indicates that the parameters were ignored in the current implementation. See f^{wave}, f^{param} in Equations (5.2),(5.7).

Condition equations

The condition equations are formed by all parameters that are regarded as observed, i.e. y_i, z_i, A, T, λ, ϕ, and follow the Equation (5.7). Every observed point i provides three condition equations — $\hat{y}_i = y_i$, $\hat{z}_i = z_i$ and Equation (5.2).

5.3.3 Method 2 — specular surface

The second developed method exploits the fact that water under some conditions can be perceived as a specularly reflecting surface. Two parallel arrays of linear lamps hang above the measurement space (see Figure 5.11). Their reflections on top of the water surface could be observed as static — in the steady wave condition, and dynamic — when the water was excited. In the latter case, the reflections project to distorted shapes (see Figure 5.7). As broadly discussed in Section 2.4, such deformations implicitly carry information about the instantaneous shape of the water surface and are investigated in the successive paragraphs.

Because the specularities (aka *highlights*) travel with the observer, no corresponding features in different cameras are present. As a result, no stereo or multi-view measurements are made possible. Unless one is able to directly identify 3D rays that would intersect at the interface of a surface (see Geometric properties in Section 2.4.1), the alternative solution to the depth-normal ambiguity is to add control information and/or impose appropriate constraints in 3D object space.

In the developed approach, the images of *specular highlights* and a number of parameter constraints are combined together to recover the water's instantaneous state. This

method solves a least squares problem, simultaneously determining all parameters of interest. The discussion opens with the condition equations and imposed constraints which constitute the functional model of the LS problem. Next, the adjustment procedure is explicitly given, including (i) the derivation of approximate values for all unknowns, (ii) the stochastic model, (iii) the system equation forming, as well as (iv) the collection of control information. The partial derivatives of constraint equations after all parameters can be found in Appedix A. Lastly, the experimental section presents the results followed by a compact conclusive paragraph.
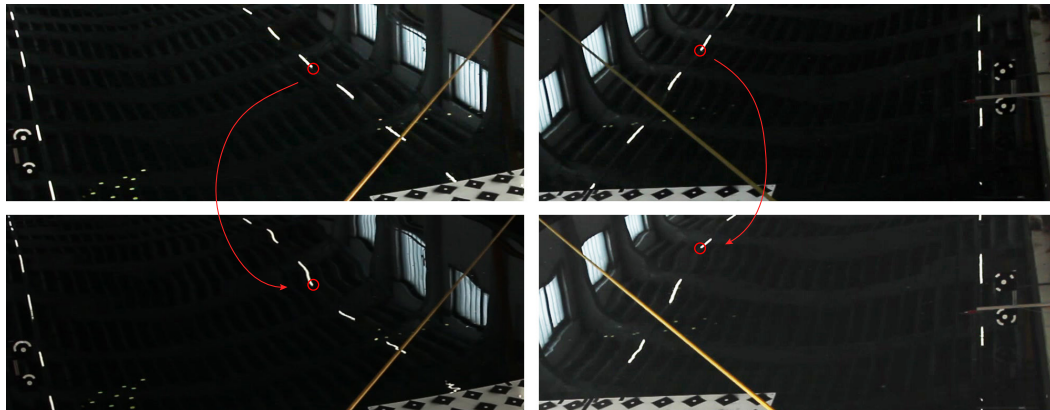

Figure 5.7: Specular reflections (*highlights*) of the control information seen by two cameras. Top: calm water condition, the water can be considered a plane, and linear features map to linear reflections. Bottom: water in excited condition, linear features map to deformed reflections.

The mathematical model

The functional model comprises the mathematical description of two observed phenomena, that is the perspective imaging associated with the camera system, as well as the shape of the induced waves, which in turn associates with a wave maker (see Figure 5.8).

The camera to object points relation was modelled with the perspective imaging equations presented in Equation (2.10). For the sake of simplicity, the exterior and interior orientation parameters as well as the additional parameters modelling the lens- and sensor-related errors were treated as constant.

The shape of the induced waves was modelled with Equation 5.2. The defined wave model is accompanied by three constraint equations. They impose that (i) virtual points lie on the wave surface (f_{surf}) (their distance from the surface =0), (ii) for all virtual points the incident and reflected rays make equal angles with respect to the tangent/normal at that point (f_{refl}) (compliance with the law of reflection), and (iii) the vector from the camera to a virtual point, the normal at the point, the vector from that

Models	Parameters
Wave	$A,\ \lambda,\ T,\ \phi,\ \mathbf{T}^{GCS\to LCS},\ \omega,\phi,\kappa^{GCS\to LCS},$ $XYZ^{virtual}$
Perspective imaging	$IOR^*,\ EOR^*,\ XYZ^{real},\ XYZ^{virtual}$

Figure 5.8: The functional model of the least squares adjustment represented in terms of the wave and the perspective imaging parameters. The wave parameters correspond to the shape defined in LCS — local coordinate system. $\mathbf{T}^{GCS\to LCS}$, $\omega,\phi,\kappa^{GCS\to LCS}$ are the parameters of a rigid transformation that bring the wave shape to GCS — global coordinate system. *denotes that in the developed implementation, parameters were treated as constant.

point towards its 3D real position are coplanar (f_{copl}) (compliance with the law of reflection). The *virtual points* are the points that correspond to the *highlights* 'moving' on top of the water surface as its shape changes. Their 3D coordinates arise from the intersections of the rays stemming from the camera with the wave model. The *real points* are the points whose reflection we observe, but in their real positions. They are always considered as ground control information, therefore the developed method belongs to the class of calibrated-environment methods. See the adjustment workflow in Figure 5.10.

The stochastic model, apart from what has been so far discussed in Subsection 5.3.2, avoids having the solution driven by the observations that are most abundant. It limits the influence of a particular group of observations with the help of a second weighting matrix N_{max}. Here, every group of observations was assigned a value n_{max} that limits its participation in the adjustment to below n_{max} observations. The diminishing effect is realised by the expression in Equation (5.5), and found on the diagonal of the matrix. The n_{obs} is the cardinality of the observations within a group. The ultimate weight matrix W' is a multiplication, $W' = W \cdot N_{max}$. The bespoke weighting strategy is implemented within *MicMac* — an open source bundle adjustment software [Deseilligny-Pierrot and Clery 2011; *MicMac, Apero, Pastis and Other Beverages in a Nutshell!* 2011].

$$n_{max}^{obs,i} = \frac{\frac{n_{obs}\,n_{max}}{n_{obs}+n_{max}}}{N_{obs}}, \quad i = 1 \dots N. \tag{5.5}$$

Within the experiments, the σ_0 value was always set to unity, whereas $\sigma_A = 1\,mm$, $\sigma_\lambda = 10\,mm$, $\sigma_T = 0.05\,s$, $\sigma_\phi = 0.25\,rad$, $\sigma_{xy} = 0.5\,pix$, $\sigma_{XYZ}^{real} = 5\,mm$, $\sigma_{XYZ}^{virtual} = 10\,mm$, $\sigma_{\mathbf{T}}^{GCS\to LCS} = 25\,mm$ and $\sigma_{\omega,\phi,\kappa}^{GCS\to LCS} = 0.01\,rad$. In analogy to *Method 1*, these values are rough estimates of the estimated approximate parameters' values. If the parameters setting shall be unclear, means of assessing the correctness of the *a priori* values must be employed, e.g. the variance component analysis.
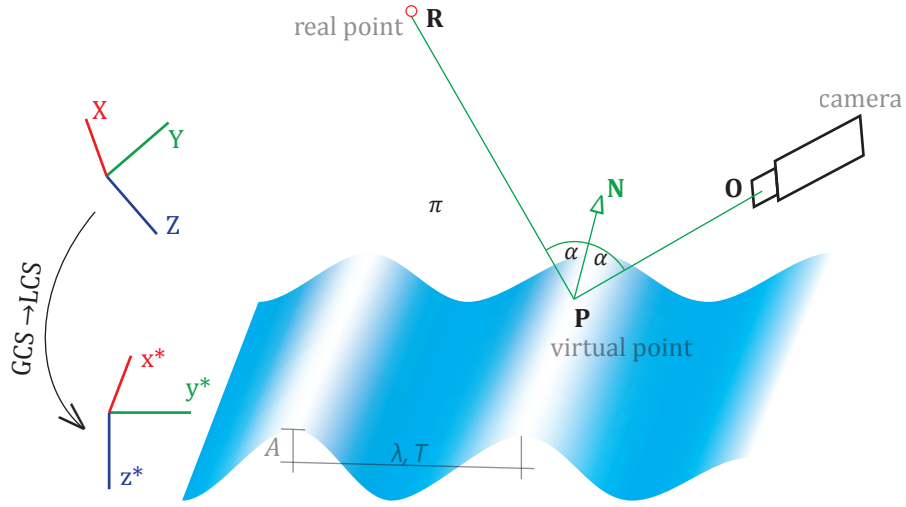
Figure 5.9: Graphical representation of the functional model. The camera captures the scene following the laws of perspective projection. The wave model, displayed at a time instance, is defined through parameters A, λ, T and ϕ in LCS. The camera (O), the virtual point **P** and real point **R** lie on a common plane π, whereas the incident and reflecting rays are symmetric w.r.t. the normal **N**.

Condition equations

Condition or observation equations are functions of observations and parameters. Collinearity equations are self-evidently observations in function of parameters — the 3D coordinates of the real or virtual point [8]. Equation (5.6) renders the collinearity expanded into Taylor series around the N initial estimates XYZ_i^0

$$\mathbf{f}_{(XYZ)}^{col} = f^{col}(XYZ_i^0) + \frac{\partial f^{col}}{\partial (XYZ)_i^0} \; d(XYZ)_i, \quad i = 1 \dots N. \tag{5.6}$$

Optionally, one may define originally free parameters as observed unknowns. This trick helps to include any available knowledge of the unknowns into the pipeline, as well as to avoid surplus parameter updates i.e. steer the rate of parameter change along the iterations. The parameters to control the rate are the entries of the weight matrix W. Our implementation allows all parameters to be regarded as observed, therefore any variable in Figure 5.8 can be replaced with the $param$ in Equation 5.7. For instance, if an X-coordinate is observed, the condition equation $\hat{X} = X^{obs} + v_x$ and the correction equation $X = X^{obs} + 1 \cdot dX$ are written down, where v_x and dX are the observation and

---

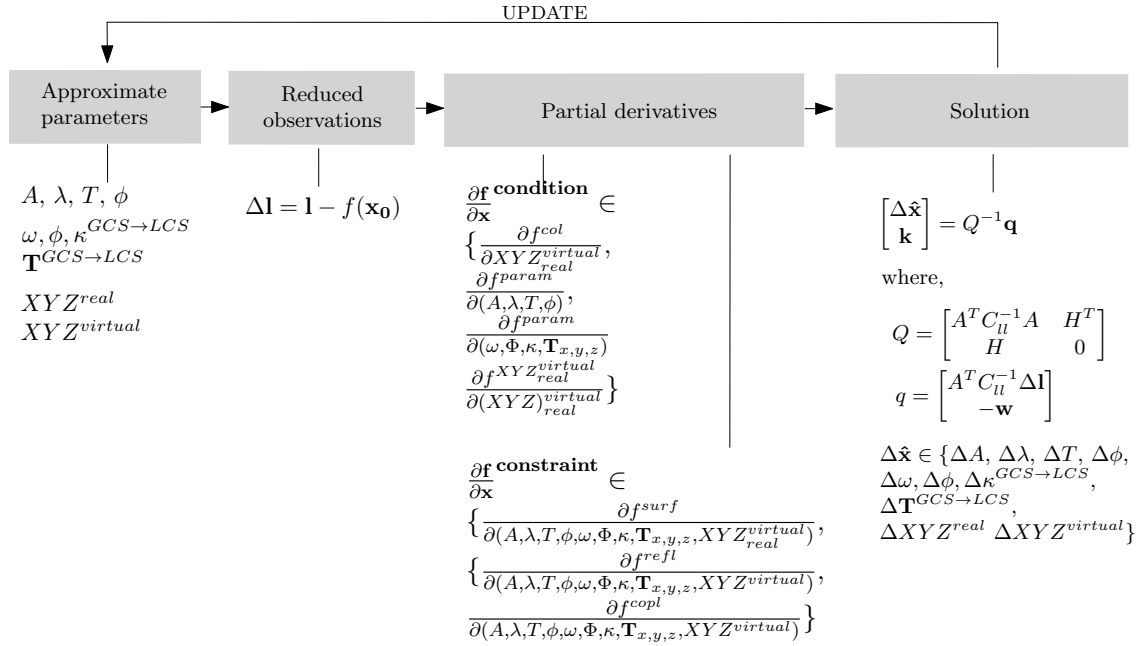[8] IOR and EOR in the developed implementation were treated as constants.

The workflow diagram at the top contains:

UPDATE

| Approximate parameters | Reduced observations | Partial derivatives | Solution |

Approximate parameters:

A, λ, T, ϕ

$\omega, \phi, \kappa^{GCS \to LCS}$

$\mathbf{T}^{GCS \to LCS}$

XYZ^{real}

$XYZ^{virtual}$

Reduced observations:

$\Delta \mathbf{l} = \mathbf{l} - f(\mathbf{x_0})$

Partial derivatives:

$\dfrac{\partial \mathbf{f}}{\partial \mathbf{x}}^{\mathbf{condition}} \in$

$\left\{ \dfrac{\partial f^{col}}{\partial XYZ^{virtual}_{real}}, \right.$

$\dfrac{\partial f^{param}}{\partial (A, \lambda, T, \phi)},$

$\dfrac{\partial f^{param}}{\partial (\omega, \Phi, \kappa, \mathbf{T}_{x,y,z})}$

$\left. \dfrac{\partial f^{XYZ^{virtual}_{real}}}{\partial (XYZ)^{virtual}_{real}} \right\}$

$\dfrac{\partial \mathbf{f}}{\partial \mathbf{x}}^{\mathbf{constraint}} \in$

$\left\{ \dfrac{\partial f^{surf}}{\partial (A, \lambda, T, \phi, \omega, \Phi, \kappa, \mathbf{T}_{x,y,z}, XYZ^{virtual}_{real})}, \right.$

$\left\{ \dfrac{\partial f^{refl}}{\partial (A, \lambda, T, \phi, \omega, \Phi, \kappa, \mathbf{T}_{x,y,z}, XYZ^{virtual})}, \right.$

$\left. \dfrac{\partial f^{copl}}{\partial (A, \lambda, T, \phi, \omega, \Phi, \kappa, \mathbf{T}_{x,y,z}, XYZ^{virtual})} \right\}$

Solution:

$\begin{bmatrix} \Delta \hat{\mathbf{x}} \\ \mathbf{k} \end{bmatrix} = Q^{-1} \mathbf{q}$

where,

$Q = \begin{bmatrix} A^T C_{ll}^{-1} A & H^T \\ H & 0 \end{bmatrix}$

$q = \begin{bmatrix} A^T C_{ll}^{-1} \Delta \mathbf{l} \\ -\mathbf{w} \end{bmatrix}$

$\Delta \hat{\mathbf{x}} \in \{\Delta A, \Delta \lambda, \Delta T, \Delta \phi, \Delta \omega, \Delta \phi, \Delta \kappa^{GCS \to LCS}, \Delta \mathbf{T}^{GCS \to LCS}, \Delta XYZ^{real} \Delta XYZ^{virtual}\}$

Figure 5.10: The least squares adjustment workflow of *Method 2*. The cycle repeats until the stop criterion is reached. $\partial \mathbf{x}$ is to denote the entire vector of parameters. See f^{col}, f^{param}, f^{surf}, f^{refl}, f^{copl} in Equations (2.10), (5.7), (5.8), (5.9), (5.10), respectively.

the parameter corrections.

$$\mathbf{f}^{param} = f^{param}_{0,i} + \frac{\partial f^{param}}{\partial (param)_i^0} \; d(param)_i, \quad i = 1 \dots N_{param}. \tag{5.7}$$

Constraint equations

Constraints do not involve observations but pure parameters. As mentioned at the beginning of the Section, the wave model renders three constraint equations
(i) $\mathbf{f}^{surf}$,

$$\mathbf{f}^{surf} = -\mathbf{z}^*(t_i) + A \; \sin\left[\, 2\pi \, \left(\frac{t_i}{T} - \frac{\mathbf{y}_i^*}{\lambda}\right) + \phi \,\right] = 0, \tag{5.8}$$

(ii) $\mathbf{f}^{refl}$,

$$\mathbf{f}^{refl} = \alpha^{OP} - \alpha^{PR} = \arccos \frac{\overrightarrow{OP}\,\overrightarrow{N}}{|\overrightarrow{OP}|\,|\overrightarrow{N}|} - \arccos \frac{\overrightarrow{PR}\,\overrightarrow{N}}{|\overrightarrow{PR}|\,|\overrightarrow{N}|} = 0 \tag{5.9}$$

(iii) $\mathbf{f}^{copl}$,

$$\mathbf{f}^{copl} = \overrightarrow{OP}\,(\overrightarrow{PR} \times \overrightarrow{N}) = \begin{vmatrix} OP_{x*} & OP_{y*} & OP_{z*} \\ RP_{x*} & RP_{y*} & RP_{z*} \\ N_{x*} & N_{x*} & N_{x*} \end{vmatrix} = 0 \qquad (5.10)$$

where $\overrightarrow{OP} = [\mathbf{x}_O^* - \mathbf{x}_P^* \quad \mathbf{y}_O^* - \mathbf{y}_P^* \quad \mathbf{z}_O^* - \mathbf{z}_P^*]$, $\overrightarrow{PR} = [\mathbf{x}_R^* - \mathbf{x}_P^* \quad \mathbf{y}_R^* - \mathbf{y}_P^* \quad \mathbf{z}_R^* - \mathbf{z}_P^*]$, $y^* = y^*(X, Y, Z, \omega, \Phi, \kappa; \mathbf{T})$ and $z^* = z^*(X, Y, Z, \omega, \Phi, \kappa; \mathbf{T})$. The constraints are defined locally, thus coordinate quantities are annexed with the symbol $*$. The values determined in LCS are not considered in the adjustment, but are obtained after a 3D rigid transformation with the parameters $(\omega, \Phi, \kappa; \mathbf{T})^{GCS \to LCS}$. f^{surf} applies to both real and virtual points, f^{refl} and f^{copl} apply solely to the latter.

The linearized forms of the above equations, expanded into Taylor series, are presented in the following. Note that the local coordinate quantities x^*, y^*, z^* are functions of their positions X, Y, Z in the GCS as well as the parameters of the 3D rigid transformation. As a result, the derivatives are calculated for a composition of functions and must obey the chain rule. The three constraint equations are then

(i) $\mathbf{f}^{surf}$,

$$\mathbf{f}^{surf} = f_{0,i}^{surf} + \frac{\partial f^{surf}}{\partial A}dA + \frac{\partial f^{surf}}{\partial T}dT + \frac{\partial f^{surf}}{\partial \lambda}d\lambda + \frac{\partial f^{surf}}{\partial \phi}d\phi + \frac{\partial f^{surf}}{\partial y_i^*}dy_i^* + \frac{\partial f^{surf}}{\partial z_i^*}dz_i^* =$$

$$f_{0,i}^{surf} + \frac{\partial f^{surf}}{\partial A}dA + \frac{\partial f^{surf}}{\partial T}dT + \frac{\partial f^{surf}}{\partial \lambda}d\lambda + \frac{\partial f^{surf}}{\partial \phi}d\phi +$$

$$\underbrace{\sum_{K \in \{X_i, Y_i, Z_i, \omega, \Phi, \kappa, \mathbf{T}^{GCS \to LCS}\}} \frac{\partial f^{surf}}{\partial y_i^*} \cdot \frac{\partial y_i^*}{\partial K}\,dK + \frac{\partial f^{surf}}{\partial z_i^*} \cdot \frac{\partial z_i^*}{\partial K}\,dK}_{\frac{\partial f^{surf}}{\partial X}, \frac{\partial f^{surf}}{\partial Y}, \frac{\partial f^{surf}}{\partial Z}, \frac{\partial f^{surf}}{\partial \omega}, \frac{\partial f^{surf}}{\partial \Phi}, \frac{\partial f^{surf}}{\partial \kappa}, \frac{\partial f^{surf}}{\partial \mathbf{T}^{GCS \to LCS}}}$$

$$(5.11)$$

(ii) $\mathbf{f}^{refl}$,

$$\mathbf{f}^{refl} = f_{0,i}^{refl} + \frac{\partial f^{refl}}{\partial x_O^*}dx_O^* + \frac{\partial f^{refl}}{\partial y_O^*}dy_O^* + \frac{\partial f^{refl}}{\partial z_O^*}dz_O^* +$$

$$\frac{\partial f^{refl}}{\partial x_R^*}dx_R^* + \frac{\partial f^{refl}}{\partial y_R^*}dy_R^* + \frac{\partial f^{refl}}{\partial z_R^*}dz_R^* +$$

$$\frac{\partial f^{refl}}{\partial N_{x*,i}}dN_{x*,i} + \frac{\partial f^{refl}}{\partial N_{y*,i}}dN_{y*,i} + \frac{\partial f^{refl}}{\partial N_{z*,i}}dN_{z*,i} =$$

$$f_{0,i}^{refl} + \sum_{j \in \{O,R\}} \sum_{K \in \{\omega, \Phi, \kappa, \mathbf{T}^{GCS \to LCS}\}} \frac{\partial f^{refl}}{\partial x_j^*} \cdot \frac{\partial x_j^*}{\partial K}dK + \frac{\partial f^{refl}}{\partial y_j^*} \cdot \frac{\partial y_j^*}{\partial K}dK + \frac{\partial f^{refl}}{\partial z_j^*} \cdot \frac{\partial z_j^*}{\partial K}dK +$$

$$\sum_{j \in \{x^*, y^*, z^*\}} \sum_{K \in \{X_i, Y_i, Z_i, \omega, \Phi, \kappa, \mathbf{T}^{GCS \to LCS}\}} \frac{\partial f^{refl}}{\partial N_{j,i}} \cdot \frac{\partial N_{j,i}}{\partial K}\,dK$$

$$(5.12)$$

(iii) $\mathbf{f}^{copl}$,

$$\mathbf{f}^{copl} = f_{0,i}^{copl} + \frac{\partial f^{copl}}{\partial x_O^*}dx_O^* + \frac{\partial f^{copl}}{\partial y_O^*}dy_O^* + \frac{\partial f^{copl}}{\partial z_O^*}dz_O^* +$$

$$\frac{\partial f^{copl}}{\partial x_R^*}dx_R^* + \frac{\partial f^{copl}}{\partial y_R^*}dy_R^* + \frac{\partial f^{copl}}{\partial z_R^*}dz_R^* +$$

$$\frac{\partial f^{copl}}{\partial N_{x^*,i}}dN_{x^*,i} + \frac{\partial f^{copl}}{\partial N_{y^*,i}}dN_{y^*,i} + \frac{\partial f^{copl}}{\partial N_{z^*,i}}dN_{z^*,i} =$$

$$f_{0,i}^{copl} + \sum_{j\in\{O,R\}} \sum_{K\in\{\omega,\Phi,\kappa,\mathbf{T}^{GCS\rightarrow LCS}\}} \frac{\partial f^{copl}}{\partial x_j^*}\cdot\frac{\partial x_j^*}{\partial K}dK + \frac{\partial f^{copl}}{\partial y_j^*}\cdot\frac{\partial y_j^*}{\partial K}dK + \frac{\partial f^{copl}}{\partial z_j^*}\cdot\frac{\partial z_j^*}{\partial K}dK +$$

$$\sum_{j\in\{x^*,y^*,z^*\}} \sum_{K\in\{X_i,Y_i,Z_i,\omega,\Phi,\kappa,\mathbf{T}^{GCS\rightarrow LCS}\}} \frac{\partial f^{copl}}{\partial N_{j,i}}\cdot\frac{\partial N_{j,i}}{\partial K}\,dK$$

(5.13)

where

$$N_{x^*,i} = \frac{\partial f^{surf}}{\partial y_i^*}\cdot\frac{\partial y_i^*}{\partial X_i} + \frac{\partial f^{surf}}{\partial z_i^*}\cdot\frac{\partial z_i^*}{\partial X_i}$$

$$N_{y^*,i} = \frac{\partial f^{surf}}{\partial y_i^*}\cdot\frac{\partial y_i^*}{\partial Y_i} + \frac{\partial f^{surf}}{\partial z_i^*}\cdot\frac{\partial z_i^*}{\partial Y_i}$$

$$N_{z^*,i} = \frac{\partial f^{surf}}{\partial y_i^*}\cdot\frac{\partial y_i^*}{\partial Z_i} + \frac{\partial f^{surf}}{\partial z_i^*}\cdot\frac{\partial z_i^*}{\partial Z_i}$$

(5.14)

Derivation of control information

The control information was not acquired physically prior to, or during the measurements. Not even posterior efforts were undertaken to collect the ground truth. The position of the linear lamps (see Figure 5.11), which served as the ground truth information, was recovered solely using the image data, under the condition that the reflecting water is globally a plane. As each measurement started with the calm water condition, the planarity condition was valid at numerous times. The imaging situation is depicted in Figure 5.12.

The calculation of the XYZ coordinates of the control information in their real locations divides into (i) the water plane derivation (from real points), (ii) identification of homologous points across views and triangulation (virtual points), lastly (iii) the flipping of the virtual points to their real positions. See Figure 5.13.

The plane π of the water was recovered thanks to well distributed dust particles present on its surface. Their appearance was sufficiently discriminative for their identification across views. Alternatively one could place artificial targets on top of the water to avoid potential identification problems. Given a few (≥ 3) pairs or triples of 2D image points corresponding to real features, their 3D position is found by intersection. The searched plane defined analytically as $Ax + By + Cz + D = 0$ is then established by *Singular Value Decomposition* (SVD).

Figure 5.11: Two arrays of lamps are marked with red bounding boxes. The end points of the single lamps served as the control information.
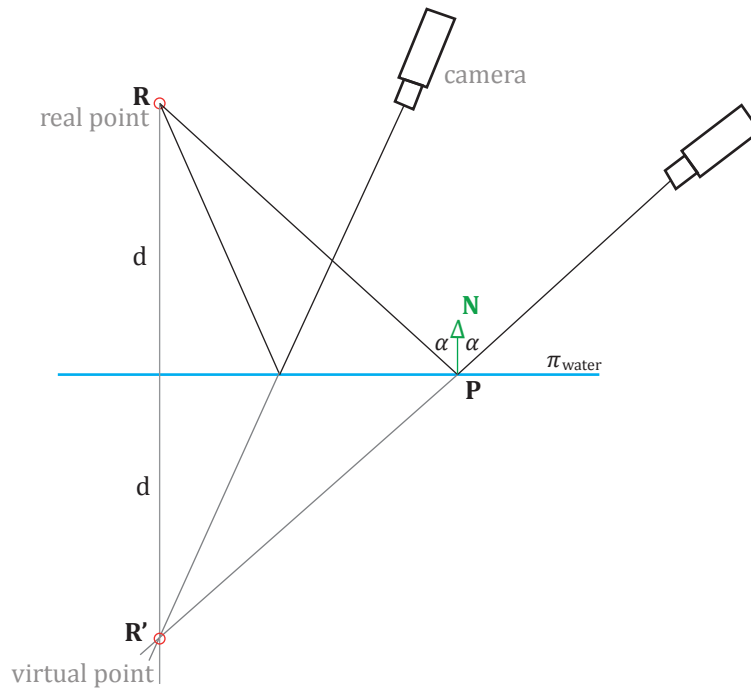

Figure 5.12: The imaging configuration taking place when deriving the control information. $\mathbf{R}$ represents the end-point of a linear lamp, $\mathbf{R}'$ is its virtual location, at d distance behind the water plane π. The camera to real-point rays follow the law of reflection i.e. the incident and reflecting rays form equal angles with the plane normal $\mathbf{N}$.

The end points of the linear lamp reflections were identified and measured in images manually only in the initial frame. The subsequent tracking in time was realised by the flagship cross-correlation technique. Having found and measured the reflections, their 3D locations are triangulated ($\mathbf{R}'$ in Figure 5.12), ignoring the fact that observed features are not real. The 3D points emerge on the wrong side of the water plane, thus must be flipped to their real positions. The flipping is done with respect to an arbitrary plane, being the water plane and determined by the coefficients A, B, C and D. The transformation performing that operation works by

(i) roto-translating the global coordinate to a local coordinate system that aligns with the flipping plane,

$$T \cdot R_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -X_i & -Y_i & -Z_i & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{\lambda}{|\mathbf{N}|} & 0 & \frac{A}{|\mathbf{N}|} & 0 \\ \frac{-AB}{|\lambda\mathbf{N}|} & \frac{C}{\lambda} & \frac{B}{|\mathbf{N}|} & 0 \\ \frac{-AC}{|\lambda\mathbf{N}|} & -\frac{B}{\lambda} & \frac{C}{|\mathbf{N}|} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.15}$$

where $\mathbf{N} = \begin{bmatrix} A & B & C \end{bmatrix}$, and $\begin{bmatrix} X_i & Y_i & Z_i \end{bmatrix}$ are 3D coordinates of any points lying within the flipping plane.

(ii) performing the actual flipping over the local $XY-$plane

$$R_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.16}$$

(iii) and bringing the point back to the global coordinate system with R_1^{-1} and T^{-1}.
The entire procedure commited in a single formula renders

$$M = T \cdot R_1 \cdot R_2 \cdot R_1^{-1} \cdot T^{-1}. \tag{5.17}$$

Derivation of approximate *highlight* position

The *highlights* coordinates ($\mathbf{P}$ in Figure 5.9) result from the intersection of the approximate wave model with the vectors anchored in the image measurements, passing through the camera perspective center and extending into the object space. The *highlights* are single image observations so there exist no correspondences across images, as in the case of real points. The intersection points are found first by intersecting with the mean water plane, and then by iteratively improving the results with the Newton's method. The points are first found in the LCS, and subsequently transferred to the GCS given the approximate parameters of the rigid transformation. The algorithm is presented below.
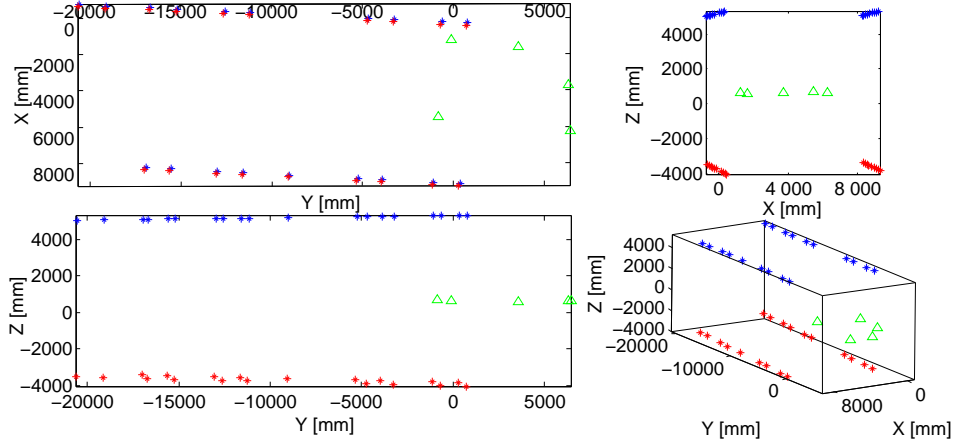
Figure 5.13: Derived control information. The points representing water surface are in green, while the blue and red correspond to the lamps in their virtual and real positions, respectively. Upper left: XY top view. Bottom left: YZ side view. Upper right: XZ side view. Bottom right: perspective view.

Given the 3D vector defined by points (x_1^*, y_1^*, z_1^*) and (x_2^*, y_2^*, z_2^*) at the camera center and observed in image space, respectively, the 3D line parametric equation takes the form

$$\begin{bmatrix} x^* \\ y^* \\ z^* \end{bmatrix} = \begin{bmatrix} x_1^* \\ y_1^* \\ z_1^* \end{bmatrix} + t \cdot \begin{bmatrix} x_2^* - x_1^* \\ y_2^* - y_1^* \\ z_2^* - z_1^* \end{bmatrix}. \tag{5.18}$$

The sought y^*−coordinate of the intersection is then equal to $y^* = y_1^* + t \cdot (y_2^* - y_1^*)$ where $t = (z^* - z_1^*)/(z_2^* - z_1^*)$. The unique solution can be obtained when the z in the preceding equation is replaced with the mean water level, e.g. $H_0 = 0$ in the LCS. A better approximation can be accomplished if the intersection is performed with a more realistic model than the plane — the observed sine wave. Combining Equation (5.2) with the last row of Equation (5.18) such that the z^* terms are equal, brings about the following relationship

$$g(y^*) = y^* - y_1^* - \frac{(y_2^* - y_1^*)}{(z_2^* - z_1^*)} \cdot (H_0 + A \, \sin[\, 2\pi \, (\frac{t}{T} - \frac{y^*}{\lambda}) + \phi \,] - z_1^*) = 0. \tag{5.19}$$

The function g has one parameter y^*, and as such, together with its derivative g' (see the derivation in Appendix A.2), both evaluated at the current parameter value y_0^*, they enter the Newton's method that finds the ultimate root. The Newton step is the ratio of $g(y_0^*)$ and $g'(y_0^*)$ (see Equation (5.20)). The loop continues until the difference between

old and new parameter estimates no longer fall below a defined threshold.

$$y^* = y_0^* - \frac{g(y_0^*)}{g'(y_0^*)} \qquad (5.20)$$

Once the y^* value is known, the z^*-coordinate is computed from Equation (5.2), and lastly the x^* can be retrieved from the 3D line equation as $x^* = x_1^* + \frac{z^* - z_1^*}{z_2^* - z_1^*} \cdot (x_2^* - x_1^*)$. The final step brings the locally determined coordinates to the global ones with $(\omega, \Phi, \kappa; \mathbf{T})^{GCS \rightarrow LCS}$.

As mentioned beforehand, in the images the highlights are tracked by the classical cross-correlation technique.

5.4 Evaluation

Results achieved with *Method 1* (*m1*) and *Method 2* (*m2*) are confronted with the responses of a Capacitive Level Sensor, and validating physical targets, see Figures 4.1, 5.14, 5.17. The Capacitive Level Sensor was mounted on a rod-like probe and sensed the variations in electrical capacity within the sensor. Given the dielectric constant of the liquid, this information can be directly transformed to the changes in the water level, in which the probe is normally immersed. Because the sensor samples the changes in a singular spot, it provides merely information on the amplitude and frequency of the water level oscillations. Likewise the validating targets. The instantaneous wavelength, thereby, remained unknown, as no direct means to judge the accuracy of the calculated wavelength existed. Indirectly, the correctness of all wave parameters, including the wavelength, can be estimated by confronting the response of a number of points distributed along the wave propagation ($vt1, vt2, cls$), with their responses predicted from the model.

Characterization of the input data is to be found in Figures 4.1, 5.15, 5.16 and 5.17 where the distribution of measured and validating points is displayed. Numerical results of the five measurement series are summarised in Figure 5.18, with the graphical representations provided in Figures 5.19, 5.20, 5.21, 5.22, 5.23, 5.24. The third and fourth measurement series was evaluated twice with varied image observations (specular highlights; cf. Figure 5.17). The adopted evaluation strategy is as follows.

Accuracy 1 — **validating targets (**$vt1$**,** $vt2$**)** With validating targets we refer to points that were not treated in the adjustments aiming at finding the wave parameters. They were measured in images and independently intersected in 3D space. The Z-response of all validating targets is confronted with the value predicted from the devised wave model. The validation takes place in the *LCS*. Figures 5.19, 5.20 illustrate the results projected onto the travelling sine wave. The red corresponds to the response from the target, the blue is the model outcome. The normal probability plots test and confirm that the residuals follow the normal distribution.

*Accuracy 2 — **the Capacitive Level Sensor (**cls**)** To compare the data collected by the Capacitive Level Sensor and the image-based measurement system, *temporal synchronisation* and *data resampling* had to take place. The start of the cls data collection was conveyed to the cameras audiovisually — by switching the light on, and by emitting a voice. It allowed for rough temporal synchronisation. To fine-align the two signals, cross-correlation was carried out. The frequency of cls data collection was double the frequency of the camera recording therefore to equalize the acquisition rates, every other sample of the cls device was discarded. Figure 5.24 illustrate the results of the comparisons.
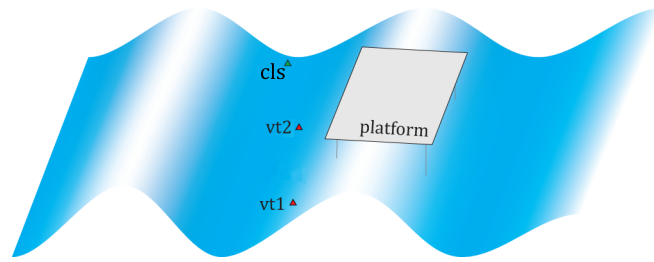
Figure 5.14: Distribution of the validating targets ($vt1$, $vt2$) and the Capacitive Level Sensor (cls) on the water surface.

ID	No. of *clusters*	No. of pairs of *clusters*	No. of conditions / parameters
$1-5_{m1}$	3	1	$2 \cdot n_{pt} \cdot n_{frame} + 5$ / 5

Figure 5.15: Characterization of input data for 5 measurement series in *Method 1*.

ID	No. of highlights	No. of GCP	No. of conditions / constraints / parameters
1_{m2}	7	4	3489 / 2091 / 2095
2_{m2}	5	4	1704 / 1020 / 1024
$3a_{m2}$	6	6	3694 / 1677 / 2017
$3b_{m2}$	6	5	4006 / 1661 / 2002
$4a_{m2}$	6	6	3377 / 1399 / 1687
$4b_{m2}$	6	4	3382 / 1402 / 1690
5_{m2}	5	4	3355 / 1575 / 1777

Figure 5.16: Characterization of input data for 5 measurement series in *Method 2*.
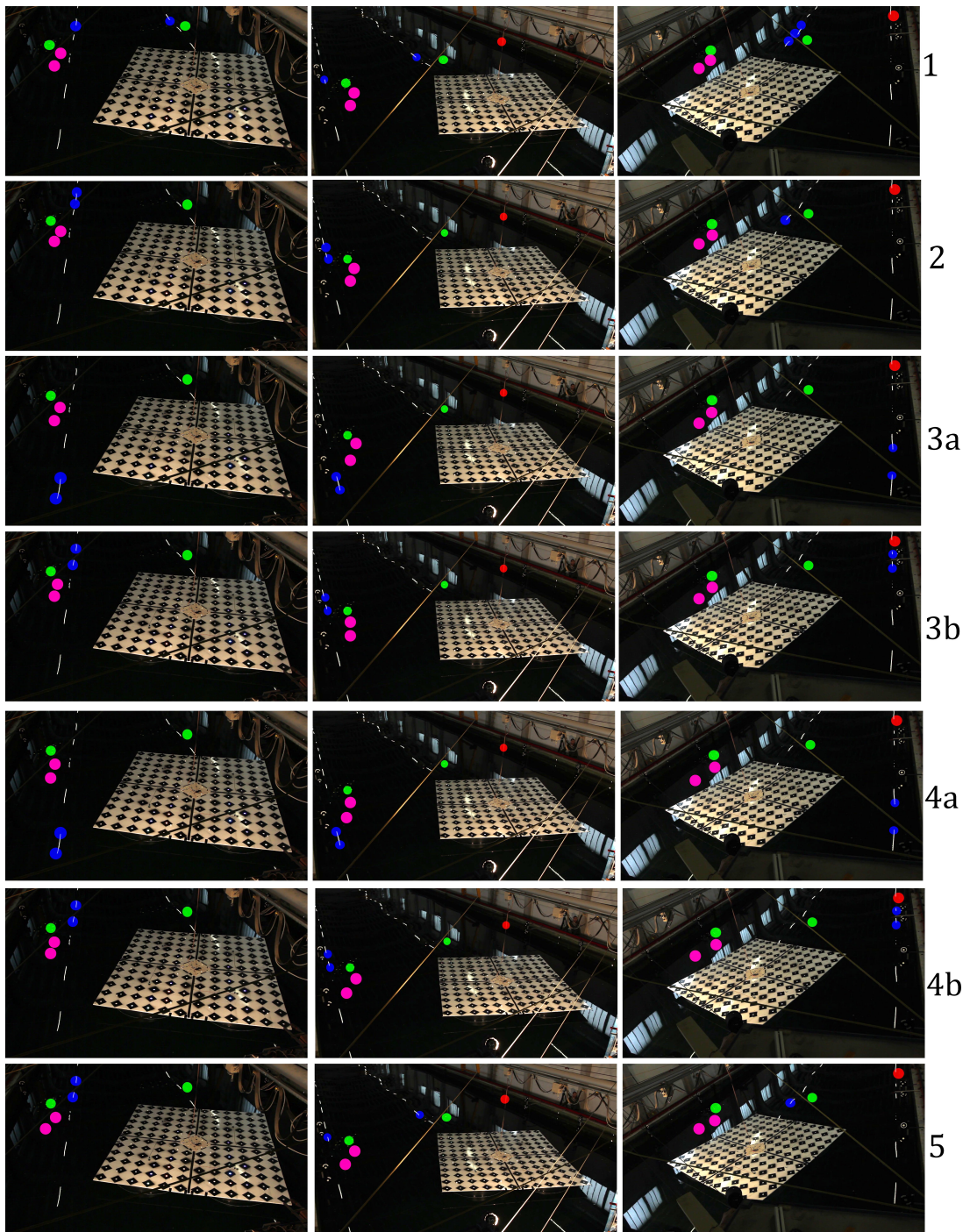
Figure 5.17: Distribution of observed highlights (blue) used in *m2*, artificial targets used in *m1* (magenta), validating targets (green) and Capacitive Level Sensor (red) in three camera views, in all measurement series.

ID	Duration [frames]	$\sigma_{x,y,z}$ [mm]	$^{(1)}RMSE_Z$ [mm] $vt1/vt2$	$^{(2)}RMSE_Z$ [mm] $vt2$
1_{m1}	70	- / 3.2 / 3.3	1.3 / 4.7	3.2
1_{m2}		1.5 / 6.5 / 0.2	1.5 / 4.1	4.0
2_{m1}	70	- / 3.3 / 3.4	2.3 / 5.5	2.3
2_{m2}		1.5 / 12 / 0.5	3.0 / 3.6	2.7
3_{m1}	83	- / 1.2 / 1.2	1.5 / 3.9	3.3
$3a_{m2}$		0.1 / 2.8 / 0.2	2.4 / 5.7	5.3
$3b_{m2}$		0.3 / 5.3 / 0.2	1.8 / 3.4	4.4
4_{m1}	70	- / 1.1 / 1.1	1.6 / 4.9	4.9
$4a_{m2}$		0.1 / 2.6 / 0.2	2.0 / 6.2	6.4
$4b_{m2}$		0.4 / 7.3 / 0.3	2.2 / 3.5	4.0
5_{m1}	100	- / 3.7 / 3.8	2.3 / 3.9	4.8
5_{m2}		1.7 / 8.7 / 0.2	2.2 / 2.8	3.7

Figure 5.18: Mean precision ($\sigma_{x,y,z}$) and accuracy ($^{(1)}RMSE_z$, $^{(2)}RMSE_z$) of 5 measurement series evaluated with the *Method 1* and *Method 2*.
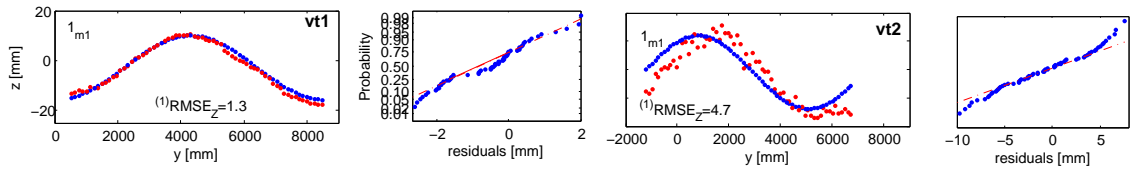
(a) $A = 12.9mm, T = 72.1, \phi = 1.0609, \lambda = 8542.8mm$
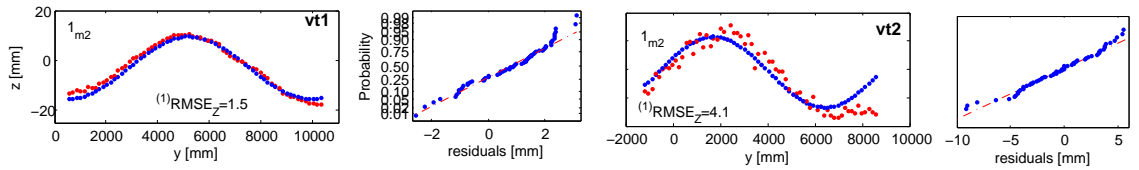
(b) $A = 12.5mm, T = 64.6, \phi = 0.7812, \lambda = 9401.6mm$

Figure 5.19: *Accuracy 1* validation results in 1st measurement series for $m1$ in (a) and $m2$ in (b), compared against the response of $vt1$ and $vt2$. Ground truth in red, fitted wave model in blue. The normal probability plots are in even columns.
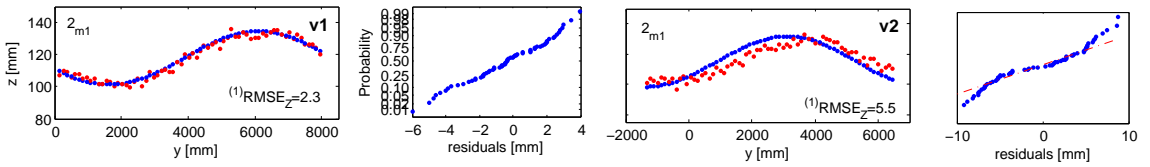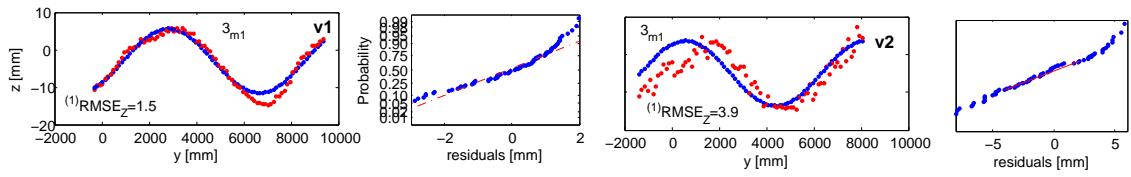

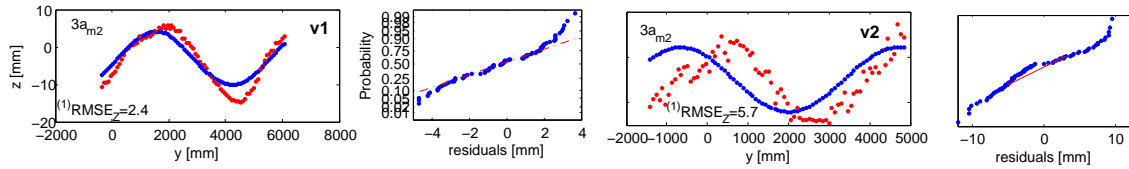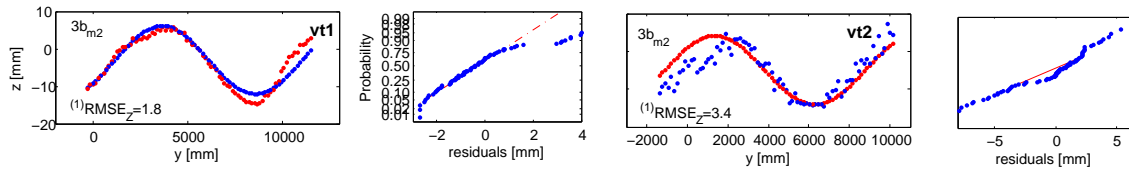(a) $A = 16.7mm, T = 76.5, \phi = -0.3932, \lambda = 8915.7mm$
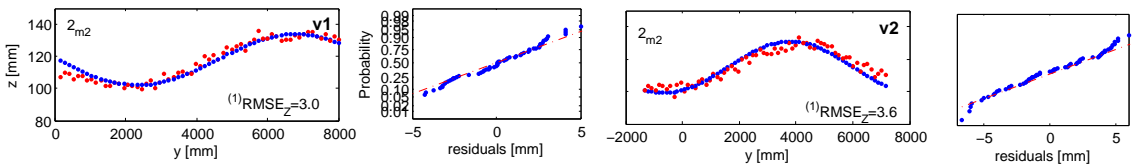
(b) $A = 15.9mm, T = 70.3, \phi = 2.3295, \lambda = 9349.4mm$

Figure 5.20: *Accuracy 1* validation results in 2nd measurement series for $m1$ in (a) and $m2$ in (b), compared against the response of $vt1$, and $vt2$. Ground truth in red, fitted wave model in blue. The normal probability plots are in even columns.

(a) $A = 8.6mm, T = 65.7, \phi = 0.0108, \lambda = 7658.4mm$

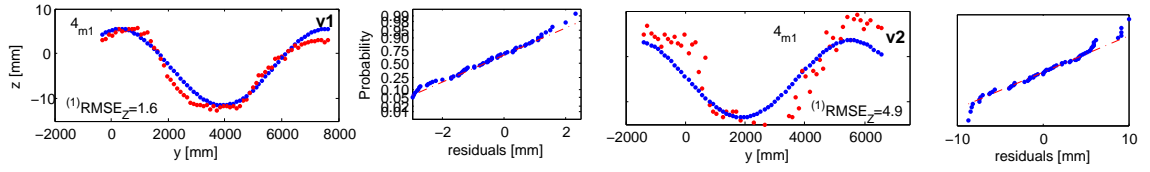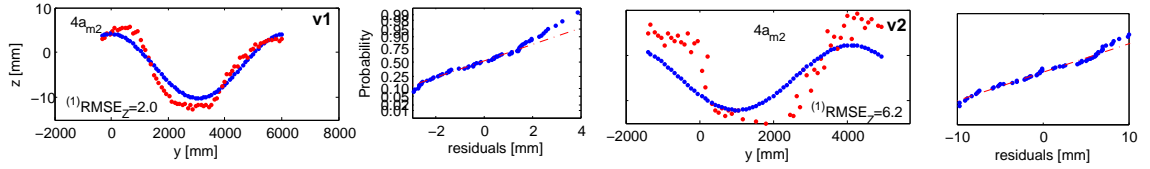(b) $A = 7.1mm, T = 69.246, \phi = 0.7794, \lambda = 5353.8.4mm$

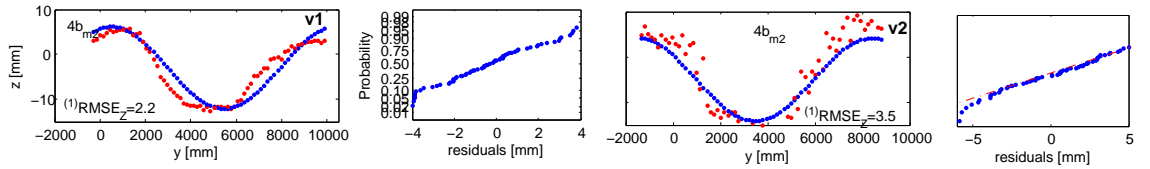(c) $A = 9.1mm, T = 69.5, \phi = 0.7756, \lambda = 9874.2mm$

Figure 5.21: *Accuracy 1* validation results in 3rd measurement series for $m1$ in (a) and $m2$ in (b)–(c), compared against the response of $vt1$, and $vt2$. Ground truth in red, fitted wave model in blue. The normal probability plots are in even columns.

(a) $A = 8.5mm, T = 63.0, \phi = -0.5305, \lambda = 7240.8mm$

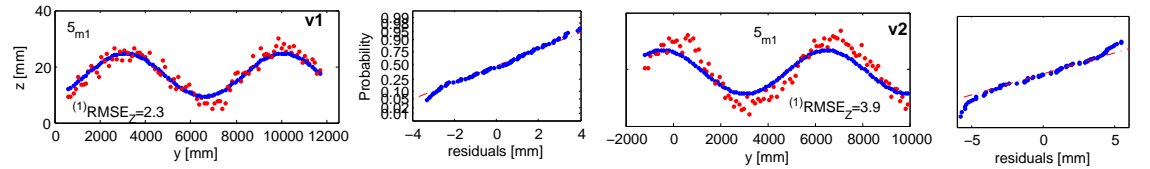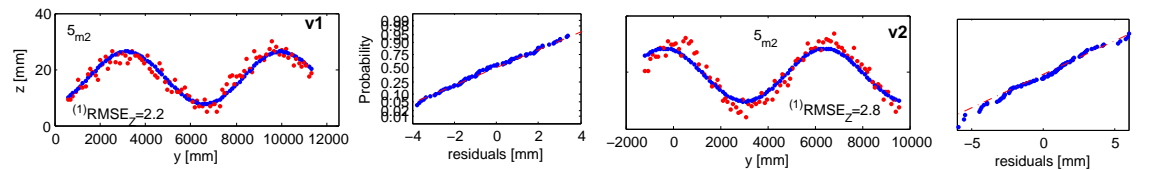(b) $A = 7.1mm, T = 70.0, \phi = 0.8025, \lambda = 6154.5mm$

(c) $A = 9.1mm, T = 70.0, \phi = 0.7821, \lambda = 9876.9mm$

Figure 5.22: *Accuracy 1* validation results in 4th measurement series for $m1$ in (a) and $m2$ in (b)–(c), compared against the response of $vt1$, and $vt2$. Ground truth in red, fitted wave model in blue. The normal probability plots are in even columns.

(a) $A = 7.7mm, T = 61.0, \phi = 2.8754, \lambda = 6900.4mm$

(b) $A = 9.3mm, T = 62.0, \phi = 0.7838, \lambda = 6775.6mm$

Figure 5.23: *Accuracy 1* validation results in 5th measurement series for $m1$ in (a) and $m2$ in (b), compared against the response of $vt1$, and $vt2$. Ground truth in red, fitted wave model in blue. The normal probability plots are in even columns.
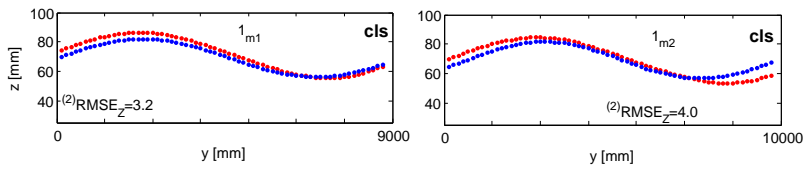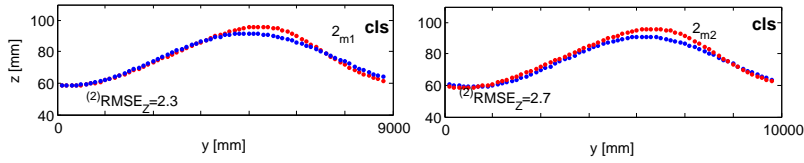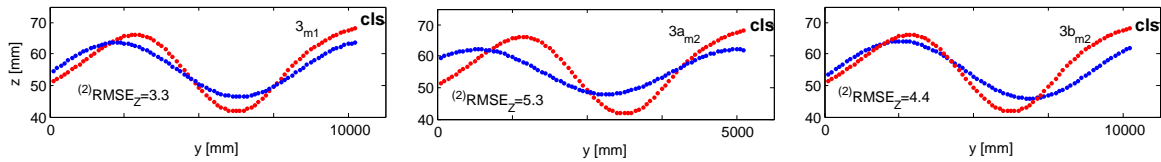
Figure 5.24: *Accuracy 2* validation results for $m1$ and $m2$ in the $1-5$ measurement series, respectively. In red the cls response, in blue the $m1$, $m2$ responses. All comparisons are carried out at position of point cls.

5.5 Discussion

The preceding sections presented two methods developed for the purpose of the water surface shape measurement. The first method regarded the water surface a diff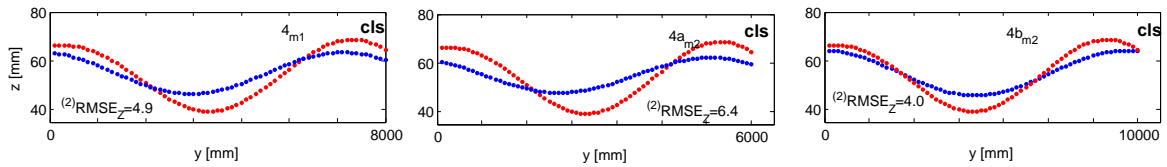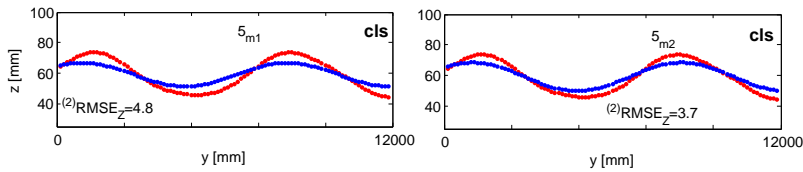use material and worked by detecting and measuring artificial targets placed on its surface, whereas the second exploited the inherent nature of the liquid medium i.e. observed water specular highlights which adapt to local surface shape. The latter required a static and calibrated illumination field outside the water basin, and can therefore be qualified as a *calibrated environment* method. In either case the modelling of the water surface was reduced to finding parameters of a travelling sine wave in a least squares adjustment.

The results achieved were confronted with responses of a Capacitive Level Sensor and two validating targets, all of which provided single point responses, and were placed in various positions across the water basin. The precision measures indicated the correctness of the performed adjustment and revealed that network design problems played a secondary role in the evaluation procedure. In an overall assessment, the *specular* method (*method 2*) proved superior with respect to the *diffuse* method (*method 1*).

Accuracy *Method 1* performs well locally, when validated on points in the vicinity of the cluster pair ($vt1$), but as soon as it is confronted with distant points ($vt2$), modelling errors significantly grow. *Method 2* employs the entire water field in the computation and therefore has a global scope with the absence of extrapolation effects, and the modelling errors more consistent yet of slightly higher magnitude. The latter was caused by the instability of water shape across the basin. The $vt1$ and cls placed on either end of the basin were under the influence of the principal sine wave, as well as the waves reflected from the basin side-walls. The platform floating in the middle also disturbed the principal wave shape. The 3rd and 4th measurement series evaluated with *method 2* on the highlights observed at the top of the basin (series $3a_{m2}$ and $4a_{m2}$) and around the platform (series $3b_{m2}$ and $4b_{m2}$) proved that the wave, having faced an obstacle, decreases its amplitude and wavelength. This is of high importance in interpreting the behaviour of the platform. To do that, one must know the form of the water body just before it hits the platform and not some distance before that interaction since it no longer corresponds with the real force put on that object.

Evaluation results on cls second the fact that the wave form changed spatio-temporally. It was systematically attenuated with increasing distance from the generating source. The cls was mounted closer to the wave maker than $vt1$, $vt2$, other artificial targets or the highlights, and consequently measured higher wave amplitudes. Wave superposition effects (the principal and reflected waves) could contribute to higher amplitudes too.

Precision Precision measures should not be interpreted as a single quality measure. As the evaluation proved, they are too optimistic when confronted with the accuracy measures. Moreover, the covariance matrices in *method 1* return standard deviation homogeneous in all coordinates, while the *method 2* manifests large uncertainty in the

106

y-coordinate. This is due to the simplified and rigorous modelling of methods 1 and 2, respectively. *Method 2* simultaneously treats the reconstruction and the modelling task, whereas *method 1* performs merely the modelling, with no special treatment of the preceding steps other than the known *a priori* standard deviation expressed in the weight matrix.

The inferior precision on the y-coordinate in *method 2* is a side effect of suboptimal network design, with a good base across the model basin (x-coordinate), and practically no shift along the y-axis (see the definition of the coordinate system in Figure 5.13).

Wave parameters   The wave parameters calculated with *method 1* and *method 2* differ, most seriously for the wavelength parameter. The differences are more pronounced at very long waves with small amplitudes (series 3, 4), and less evident at shorter waves or high amplitudes (series 1, 2, 5). In cases of small amplitude to wavelength ratios, stability of the solution ought to be brought into discussion. Despite of that, the subject has not been given further insight within this work.

CHAPTER 6

Conclusion

The goal of the research was to build an imaging system consisting of non-professional cameras, supply it with suitable algorithms and examine its performance in terms of

- feasibility of observing timely phenomena,

- feasibility of observing objects with difficult surfaces, and

- achievable accuracy.

This was accomplished by employing the setup for a surveying task in an industrial application. A (diffuse) platform floating on the water and the surrounding (specular) water surface were simultaneously analysed. The motion component came from the water wave excitation. The results were confronted with independent measurements acquired by a motion capture system (MCS). Overall, the thesis demonstrated the plausibility of using a low-budget imaging set-up to serve for precise metrology purposes.

Hardware (Subsections 4.3.1, 4.3.2)   The system built from three dSLR cameras was far from being a turnkey system. The predicaments originated from e.g. lack of means to synchronise multi-camera acquisitions. Data collection from each camera therefore occurred sequentially and required that post-processing delay corrections were applied. If longer acquisitions are in mind, the lack of an external power supply appears as an additional hurdle. Exchanging the batteries introduces forces on the camera body, possibly causing unstable camera orientations, which prompts re-calibration. No such issues were experienced, but were they a case, best advice would be to perform a series of camera calibration throughout the acquisition. If, for instance, only two calibrations were performed, the camera parameters between the two instances can be retrieved by interpolation. The author does not believe that consumer grade camera producers will ever replace with or offer in addition to the battery packs, also the external power supplies.

Unlike the low-cost solutions, industrial cameras allow for the current flow through USB, the Ethernet or some hybrid cables.

The MCS was certainly more universal, in the sense that it could be useful in different applications at no extra effort, it was, however, less adaptable. On the contrary, our system was less universal but more task-specific. Because we were responsible for constructing both the hardware and software, our system offered more flexibility in how and what to survey. Unlike the maximum 16 targets of the MCS, we were free to discretize our object by any number of targets. As was proved, with the same equipment we were even able to observe and measure the specular reflections on the water surface.

The temporal resolution of our dSLR cameras, however, was always inferior to that of MCS. The limits of a dSLR camera are 30 frames per second for the Full-HD, and 60 frames per second at a reduced spatial resolution. In this respect, one is restrained to resolve phenomena at a frequency of less than 15Hz and 30Hz, respectively. Potential future investigations could see whether the today's 4K UHD sensors, the backside illumination sensors (BI-CMOS), and higher frame rates place one closer to the MCSs. It is not yet evident if the new technology is indeed an advancement. As it is predominantly present among the digital single lens mirrorless cameras (dSLMs), the bodies are attached to a single, universal zoom lens — and moreover — they are said to suffer from the rolling shutter which could be perceived a serious downside.

The consequence of the MCSs wide range of temporal resolutions is that the processing, i.e. triangulation to points, takes place on-board, and only the 3D coordinates are sent to the PC. The image information is lost. On the other hand, our system acquired in a less rapid fashion but each image frame was stored and could be reviewed *a posteriori* if needed. Consequently, if image points were incorrectly measured one could track back to the source of the error and have the opportunity to re-do the point localization.

It is also worth highlighting the distinction between the wavelength sensitivity of the two sensors. A dSLR camera captures color via a color filter, typically the Bayer filter. On top of every pixel a filter is placed which lets only part of the signal through (red, green or blue), and discards the rest. Therefore, one needs to integrate the incoming signal over the given pixel for longer times than would be necessary with a monochrome sensor specific to MCSs. Put differently, for equal rate of acquisition the MCS will demand less illumination on the object than the dSLR-composed measurement setup.

Shape reconstruction in time (Subsections 4.3.3, 4.3.4)   The 3D shape reconstruction took place in both static and dynamic conditions with the cameras still and mobile, respectively. All in all, the assembled hardware produced satisfying outcomes, consistent with the ground truth provided by the MCS. At an average image scale of 1:500 the precision and accuracy figures oscillated around $2mm$.

Several hardships and restrictions arose at the stage of point detection and tracking, and were inherent of either the sensor characteristic (see previous paragraph) or the workplace environment. Firstly, the lossy compression of the videos contributed to block artefacts, thereby, increased noise levels within the image measurements. Secondly, occlusions from neighbouring objects and insufficient illumination led to timely loss of

observed points. It is highly advisable to avoid such circumstances yet when faced, they can be potentially overcome with the help of predicting tools, e.g. Kalman filter. Proper project planning must also be paid close attention to. Activities undertaken prior to the survey, such as camera distribution, visibility analyses or target design must be well adapted to the image scale and resolution, illumination power and the ambient surrounding itself. Only proceeding this way guarantees optimal results.

Specular surface shape reconstruction (Chapter 5) Measurement of the difficult surface was approached with the collinearity equation — treating it as diffuse, and with the piecewise linear equations — when restituted solely from the specular reflections visible on its surface but coming from a set of lamps hung from the ceiling. The accuracies obtained on physical targets floating on the surface, counted for the entire acquisition length, were more favourable for the latter method, falling between 1 and $3mm$.

The concept of using the lamps' reflections for metrology was partially driven by the fact that lamps in ship testing facilities appear predominantly in the same configurations. They provide a calibration field at no labour cost, of high reliability, making the methodology universal and easy to re-apply. The *diffuse* method, on the contrary, necessitates extra work to establish a net of targets to be placed on the water. The points are then prone to sinking, partial submergence, occlusions or drift.

The superiority of *specular* over *diffuse* is in full-field versus single point shape modelling. To install a net of points that spans a large area is infeasible, therefore one is constrained to local observations. On the contrary, the number of specular highlights is a multiplication of every point in the calibration field by the number of cameras within the imaging system. Their allocation over the measurement volume is steerable by the camera placement. If, however, bound to using the diffuse approach and aiming at full-field data collection, eliciting the water shape over extended surfaces may be possible through adoption of patches of nets in strategic areas. In both conditions, large field modelling demands very careful planning, especially for complex-shape surfaces. The model definition must contain just enough parameters, whereas the observations ought to deliver enough data for their recovery.

A noteworthy aspect of the *specular* approach is the magnifying effect present on the water surface. Observed highlights undergo an apparent motion under the deforming surface shape. The motion magnitude and trajectory are known from the law of reflection. It depends on the camera to surface relation, and the relation of the point of which reflection is observed (in its real position) to its reflection on the surface (virtual position). By modifying the distance between the reflecting surface and the calibration field (real points), the motion magnitude is changed proportionally. Put differently, very small surface deformations can render large highlight displacements for a sufficiently distant calibration field.

An important issue to consider when doing least squares adjustment — true for the *diffuse* and *specular* methods — is the initial approximations of all unknowns. Unless their values are known well enough, the success of the adjustment is put under question. At small amplitudes, the longer the wave is, the more precise must be the approx-

imations. If approximations are imprecise, divergence or convergence to an incorrect solution is highly probable. Reliability measures output from the covariance matrices of the adjustment may serve to evaluate the credibility of the results. Nonetheless, this has not been investigated within this thesis.

The model of the wave shape assumes single frequency oscillations. In large field observations, this assumption is often violated, as has been observed in the presented work. If the assumption is violated, and the model becomes insufficient to describe the phenomena, one may still (i) use the simple model to observe the surface locally, or (ii) extend it to involve wave time-varying components, eventually model its shape with a wave being a sum of two elementary waves.

Partial derivatives

A.1   Constraint equations

This Section presents partial derivatives that occur in the linearized Equation (5.11). The derivatives of Equations (5.12) and (5.13) are not explicated hereafter due to their verbose form. They can be easily found using automated derivatives retrieval techniques, such that the one present in the Matlab Symbolic Toolbox.

The coordinates subscripted with P refer to the virtual location of the real point, all three lying on the wave surface and given in GCS; see also Figure 5.9 for a clearer picture. Note the following substitutions, $\mathrm{sn} = \sin$, $\mathrm{cs} = \cos$, $\Delta X = X_P - T_x$, $\Delta Y = Y_P - T_y$.

$$\frac{\partial f^{surf}}{\partial A} = \mathrm{sn}\left(\phi - 2\pi \cdot \left(\frac{y^*}{\lambda} + \frac{t}{T}\right)\right), \qquad \frac{\partial f^{surf}}{\partial T} = \frac{-2\pi \cdot A \cdot t \cdot \mathrm{cs}\left(\phi - 2\pi\left(\frac{y^*}{\lambda} + \frac{t}{T}\right)\right)}{T^2},$$

$$\frac{\partial f^{surf}}{\partial \lambda} = \frac{2\pi \cdot A \cdot \mathrm{cs}\left(\phi - 2\pi \cdot \left(\frac{y^*}{\lambda} + \frac{t}{T}\right) \cdot y^*\right.}{\lambda^2} \qquad \frac{\partial f^{surf}}{\partial \phi} = A \cdot \mathrm{cs}\left(\phi - 2\pi \cdot \left(\frac{y^*}{\lambda} + \frac{t}{T}\right)\right),$$

$$\frac{\partial f^{surf}}{\partial \omega} = \frac{\partial f^{surf}}{\partial y^*} \cdot \frac{\partial y^*}{\partial \omega} + \frac{\partial f^{surf}}{\partial z^*} \cdot \frac{\partial z^*}{\partial \omega} = \mathrm{cs}\omega \cdot \mathrm{cs}\Phi \cdot \Delta Y + \mathrm{cs}\Phi \cdot \mathrm{sn}\omega \cdot \Delta Z +$$
$$2\pi \cdot A \cdot \mathrm{cs}(\phi - 2\pi \cdot ((\mathrm{cs}\kappa \cdot \mathrm{cs}\omega - \mathrm{sn}\kappa \cdot \mathrm{sn}\omega \cdot \mathrm{sn}\Phi) \cdot \Delta Y + (\mathrm{cs}\kappa \cdot \mathrm{sn}\omega + \mathrm{cs}\omega \cdot \mathrm{sn}\kappa \cdot \mathrm{sn}\Phi) \cdot \Delta Z -$$
$$\mathrm{cs}\Phi \cdot \mathrm{sn}\kappa \cdot \Delta X)/\lambda + 2\pi t/T) \cdot ((\mathrm{cs}\kappa \cdot \mathrm{sn}\omega + \mathrm{cs}\omega \cdot \mathrm{sn}\kappa \cdot \mathrm{sn}\Phi) \cdot \Delta Y -$$
$$(\mathrm{cs}\kappa \cdot \mathrm{cs}\omega - \mathrm{sn}\kappa \cdot \mathrm{sn}\omega \cdot \mathrm{sn}\Phi) \cdot \Delta Z)/\lambda,$$

$$\frac{\partial f^{surf}}{\partial \Phi} = \frac{\partial f^{surf}}{\partial y^*} \cdot \frac{\partial y^*}{\partial \Phi} + \frac{\partial f^{surf}}{\partial z^*} \cdot \frac{\partial z^*}{\partial \Phi} = \mathrm{cs}\omega \cdot \mathrm{sn}\Phi \cdot \Delta Z - \mathrm{cs}\Phi \cdot \Delta X - \mathrm{sn}\omega \cdot \mathrm{sn}\Phi \cdot \Delta Y -$$
$$2\pi \cdot A \cdot \mathrm{cs}(\phi - 2\pi \cdot ((\mathrm{cs}\kappa \cdot \mathrm{cs}\omega - \mathrm{sn}\kappa \cdot \mathrm{sn}\omega \cdot \mathrm{sn}) \cdot \Delta Y + (\mathrm{cs}\kappa \cdot \mathrm{sn}\omega + \mathrm{cs}\omega \cdot \mathrm{sn}\kappa \cdot \mathrm{sn}\Phi) \cdot \Delta Z -$$
$$\mathrm{cs}\Phi \cdot \mathrm{sn}\kappa \cdot \Delta X)/\lambda + 2\pi t/T) \cdot (\mathrm{sn}\kappa \cdot \mathrm{sn}\Phi \cdot \Delta X + \mathrm{cs}\omega \cdot \mathrm{cs}\Phi \cdot \mathrm{sn}\kappa \cdot \Delta Z - \mathrm{cs}\Phi \cdot \mathrm{sn}\kappa \cdot \mathrm{sn}\omega \cdot \Delta Y)/\lambda$$

$$\frac{\partial f^{surf}}{\partial \kappa} = \frac{\partial f^{surf}}{\partial y^*} \cdot \frac{\partial y^*}{\partial \kappa} + \frac{\partial f^{surf}}{\partial z^*} \cdot \frac{\partial z^*}{\partial \kappa} = 2\pi A \cdot \mathrm{cs}(\phi - 2\pi((\mathrm{cs}\kappa \cdot \mathrm{cs}\omega - \mathrm{sn}\kappa \cdot \mathrm{sn}\omega \cdot \mathrm{sn}\Phi)\Delta Y +$$
$$(\mathrm{cs}\kappa \cdot \mathrm{sn}\omega + \mathrm{cs}\omega \cdot \mathrm{sn}\kappa \cdot \mathrm{sn}\Phi)\Delta Z - \mathrm{cs}\Phi \cdot \mathrm{sn}\kappa \Delta X)/\lambda + 2\pi t/T) \cdot$$
$$((\mathrm{cs}\omega \cdot \mathrm{sn}\kappa + \mathrm{cs}\kappa \cdot \mathrm{sn}\omega \cdot \mathrm{sn}\Phi) \cdot \Delta Y + (\mathrm{sn}\kappa \cdot \mathrm{sn}\omega - \mathrm{cs}\kappa \cdot \mathrm{cs}\omega \cdot \mathrm{sn}\Phi)\Delta Z + \mathrm{cs}\kappa \cdot \mathrm{cs}\Phi \Delta X)/\lambda$$

$$\frac{\partial f^{surf}}{\partial T_x} = \frac{\partial f^{surf}}{\partial y^*} \cdot \frac{\partial y^*}{\partial T_x} + \frac{\partial f^{surf}}{\partial z^*} \cdot \frac{\partial z^*}{\partial T_x} = \mathrm{sn}\Phi - (2\pi A \cdot \mathrm{cs}(\phi - 2\pi((\mathrm{cs}\kappa \cdot \mathrm{cs}\omega - \mathrm{sn}\kappa \cdot \mathrm{sn}\omega \cdot \mathrm{sn}\Phi) \cdot$$
$$\Delta Y + (\mathrm{cs}\kappa \cdot \mathrm{sn}\omega + \mathrm{cs}\omega \cdot \mathrm{sn}\kappa \cdot \mathrm{sn}\Phi)\Delta Z - \mathrm{cs}\Phi \cdot \mathrm{sn}\kappa \cdot \Delta X)/\lambda + 2\pi t/T) \cdot \mathrm{cs}\Phi \cdot \mathrm{sn}\kappa)/\lambda$$

$$\frac{\partial f^{surf}}{\partial T_y} = \frac{\partial f^{surf}}{\partial y^*} \cdot \frac{\partial y^*}{\partial T_y} + \frac{\partial f^{surf}}{\partial z^*} \cdot \frac{\partial z^*}{\partial T_y} = (2\pi A \cdot \mathrm{cs}(\phi - (2\pi((\mathrm{cs}\kappa \cdot \mathrm{cs}\omega - \mathrm{sn}\kappa \cdot \mathrm{sn}\omega \cdot \mathrm{sn}\Phi)\Delta Y +$$
$$(\mathrm{cs}\kappa \cdot \mathrm{sn}\omega + \mathrm{cs}\omega \cdot \mathrm{sn}\kappa \cdot \mathrm{sn}\Phi)\Delta Z - \mathrm{cs}\Phi \cdot \mathrm{sn}\kappa \cdot \Delta X))/\lambda + 2\pi t/T) \cdot$$
$$(\mathrm{cs}\kappa \cdot \mathrm{cs}\omega - \mathrm{sn}\kappa \cdot \mathrm{sn}\omega \cdot \mathrm{sn}\Phi))/\lambda - \mathrm{cs}\Phi \cdot \mathrm{sn}\omega$$

$$\frac{\partial f^{surf}}{\partial T_z} = \frac{\partial f^{surf}}{\partial y^*} \cdot \frac{\partial y^*}{\partial T_z} + \frac{\partial f^{surf}}{\partial z^*} \cdot \frac{\partial z^*}{\partial T_z} = \mathrm{cs}\omega \cdot \mathrm{cs}\Phi +$$
$$(2\pi A \cdot \mathrm{cs}(\phi - (2\pi((\mathrm{cs}\kappa \cdot \mathrm{cs}\omega - \mathrm{sn}\kappa \cdot \mathrm{sn}\omega \cdot \mathrm{sn}\Phi)\Delta Y + (\mathrm{cs}\kappa \cdot \mathrm{sn}\omega + \mathrm{cs}\omega \cdot \mathrm{sn}\kappa \cdot \mathrm{sn}\Phi)\Delta Z -$$
$$\mathrm{cs}\Phi \cdot \mathrm{sn}\kappa \cdot \Delta X))/\lambda + 2\pi t/T) \cdot (\mathrm{cs}\kappa \cdot \mathrm{sn}\omega + \mathrm{cs}\omega \cdot \mathrm{sn}\kappa \cdot \mathrm{sn}\Phi))/\lambda$$

$$\frac{\partial f^{surf}}{\partial X} = \frac{\partial f^{surf}}{\partial y^*} \cdot \frac{\partial y^*}{\partial X} + \frac{\partial f^{surf}}{\partial z^*} \cdot \frac{\partial z^*}{\partial X} = (2\pi A \cdot \mathrm{cs}(\phi - 2\pi((\mathrm{cs}\kappa \cdot \mathrm{cs}\omega - \mathrm{sn}\kappa \cdot \mathrm{sn}\omega \cdot \mathrm{sn}\Phi) \cdot \Delta Y +$$
$$(\mathrm{cs}\kappa \cdot \mathrm{sn}\omega + \mathrm{cs}\omega \cdot \mathrm{sn}\kappa * \mathrm{sn}\Phi) \cdot \Delta Z - \mathrm{cs}\Phi \cdot \mathrm{sn}\kappa \cdot \Delta X)/\lambda + 2\pi t/T) \cdot \mathrm{cs}\Phi \cdot \mathrm{sn}\kappa)/\lambda - \mathrm{sn}\Phi$$

$$\frac{\partial f^{surf}}{\partial Y} = \frac{\partial f^{surf}}{\partial y^*} \cdot \frac{\partial y^*}{\partial Y} + \frac{\partial f^{surf}}{\partial z^*} \cdot \frac{\partial z^*}{\partial Y} = \mathrm{cs}\Phi \cdot \mathrm{sn}\omega -$$
$$(2\pi A \cdot \mathrm{cs}(\phi - 2\pi((\mathrm{cs}\kappa \cdot \mathrm{cs}\omega - \mathrm{sn}\kappa \cdot \mathrm{sn}\omega \cdot \mathrm{sn}\Phi)\Delta Y + (\mathrm{cs}\kappa \cdot \mathrm{sn}\omega + \mathrm{cs}\omega \cdot \mathrm{sn}\kappa \cdot \mathrm{sn}\Phi)\Delta Z -$$
$$\mathrm{cs}\Phi \cdot \mathrm{sn}\kappa \cdot \Delta X)/\lambda + 2\pi t/T) \cdot (\mathrm{cs}\kappa \cdot \mathrm{cs}\omega - \mathrm{sn}\kappa \cdot \mathrm{sn}\omega \cdot \mathrm{sn}\Phi))/\lambda$$

$$\frac{\partial f^{surf}}{\partial Z} = \frac{\partial f^{surf}}{\partial y^*} \cdot \frac{\partial y^*}{\partial Z} + \frac{\partial f^{surf}}{\partial z^*} \cdot \frac{\partial z^*}{\partial Z} = -\mathrm{cs}\omega \cdot \mathrm{cs}\Phi -$$
$$(2\pi A \cdot \mathrm{cs}(\phi - 2\pi((\mathrm{cs}\kappa \cdot \mathrm{cs}\omega - \mathrm{sn}\kappa \cdot \mathrm{sn}\omega \cdot \mathrm{sn}\Phi)\Delta Y + (\mathrm{cs}\kappa \cdot \mathrm{sn}\omega + \mathrm{cs}\omega \cdot \mathrm{sn}\kappa \cdot \mathrm{sn}\Phi)\Delta Z -$$
$$\mathrm{cs}\Phi \cdot \mathrm{sn}\kappa \cdot \Delta X)/\lambda + 2\pi t/T) \cdot (\mathrm{cs}\kappa \cdot \mathrm{sn}\omega + \mathrm{cs}\omega \cdot \mathrm{sn}\kappa \cdot \mathrm{sn}\Phi))/\lambda$$

A.2 3D line — sinusoid intersection

Given below is the first order derivative of function in Equation (5.19) after parameter y.

$$g'(y) = \frac{2\pi \cdot A \cdot \cos(\phi + 2\pi \cdot (\frac{t}{T} - \frac{y}{\lambda})) \cdot (y_1 - y_2)}{\lambda \cdot (z_1 - z_2) + 1} \tag{A.1}$$

Bibliography

Aarninkhof, S. G., Turner, I. L., Dronkers, T. D., Caljouw, M., and Nipius, L. [2003]. "A video-based technique for mapping intertidal beach bathymetry". In: *Coastal Engineering* 49 (4), pp. 275–289.

Abdel-Aziz, Y. and Karara, H. [1971]. "Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry". In: *ASP Symp. on Close Range Photogrammetry*. University of Illinois. Urbana, pp. 1–18.

Adams, L. and Pos, J. [1984]. "Wave height measurements in model harbours using close range photogrammetry". In: *15th Congress of the International Society for Photogrammetry and Remote Sensing, Rio de Janeiro*.

Adato, Y., Vasilyev, Y., Ben-Shahar, O., and Zickler, T. [2007]. "Toward a theory of shape from specular flow". In: *IEEE 11th International Conference on Computer Vision. ICCV.* Pp. 1–8.

Arie-Nachimson, M., Kovalsky, S. Z., Kemelmacher-Shlizerman, I., Singer, A., and Basri, R. [2012]. "Global motion estimation from point matches". In: *International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*. Vol. 2. IEEE, pp. 81–88.

Arrigonis, F., Rossi, B., Malapelle, F., Fragneto, P., and Fusiello, A. [2014]. "Robust global motion estimation with matrix completion". In: *ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 45, pp. 63–70.

Benning, W, Lange, J, Schwermann, R, Effkemann, C, and Görtz, S [2004]. "Monitoring crack origin and evolution at concrete elements using photogrammetry". In: *ISPRS Congress Istanbul Commission*.

Bhat, D. N. and Nayar, S. K. [1995]. "Stereo in the Presence of Specular Reflection". In: *Proceedings of the 5th International Conference on Computer Vision (ICCV)*. Pp. 1086–1092.

Bhat, D. N. and Nayar, S. K. [1998]. "Stereo and specular reflection". In: *International Journal of Computer Vision* 26 (2), pp. 91–106.

Black, J. T., Blandino, J. R., Jones, T. W., Danehy, P. M., and Dorrington, A. A. [2003]. *Dot-Projection Photogrammetry and Videogrammetry of Gossamer Space Structures*. Tech. rep. NASA/TM-2003-212146. NASA Langley Research Center.

Blake, A. [1985]. "Specular Stereo". In: *Proceedings of the 9th International Joint Conference on Artificial Intelligence*. Vol. 2. San Francisco, CA, USA, pp. 973–976.

Blake, A. and Brelstaff, G. [1988]. "Geometry From Specularities." In: *International Conference on Computer Vision (ICCV)*, pp. 394–403.

Blandino, J. R., Sterling, J., Baginski, F., Steadman, E., Black, J. T., and Pappa, R. S. [2004]. "Optical strain measurement of an inflated cylinder using photogrammetry with application to scientific balloons". In: *Proceedings of Collection of Technical PapersAIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, pp. 1–13.

Bock, E. J. and Hara, T. [1995]. "Optical measurements of capillary-gravity wave spectra using a scanning laser slope gauge". In: *Journal of Atmospheric and Oceanic Technology* 12 (2), pp. 395–403.

Bonfort, T. and Sturm, P. [2003]. "Voxel carving for specular surfaces". In: *Proceedings of the 9th International Conference on Computer Vision (ICCV)*. Pp. 591–596.

Börlin, N. and Grussenmeyer, P. [2013]. "Bundle adjustment with and without damping". In: *The Photogrammetric Record* 28 (144), pp. 396–415.

Bösemann, W. [1996]. "The Optical Tube Measurement System OLM Photogrammetric Methods used for Industrial Automation and Process Control". In: *ISPRS International Archives of Photogrammetry and Remote Sensing* 31, pp. 55–58.

— [2011]. "Industrial photogrammetry: challenges and opportunities". In: *Proc. SPIE, Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection*, 80850H–80850H.

Brown, D. C. [1956]. *The simultaneous determination of the orientation and lens distortion of a photogrammetric camera*. Tech. rep. 58-8. Florida: Patrick AFB: Air Force Missile Test Center.

— [1971]. "Close-range camera calibration". In: *Photogrammetric Engineering* 37, pp. 855–866.

Burgess, G, Shortis, M., and Scott, P [2011]. "Photographic assessment of retroreflective film properties". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (5), pp. 743–750.

Chandler, J. H., Ferreira, E., Wackrow, R., and Shiono, K. [2014]. "Water Surface and Velocity Measurement-River and Flume". In: *ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 45, pp. 151–156.

Chandler, J., Wackrow, R, Sun, X, Shiono, K, and Rameshwaran, P [2008]. "Measuring a dynamic and flooding river surface by close range digital photogrammetry". In: *ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (B7), pp. 211–216.

Chen, T., Goesele, M., and Seidel, H.-P. [2006]. "Mesostructure from specularity". In: *Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2, pp. 1825–1832.

Clarke, T. and Fryer, J. [1998]. "The development of camera calibration methods and models". In: *The Photogrammetric Record* 16 (91), pp. 51–66.

Cobelli, P. J., Maurel, A., Pagneux, V., and Petitjeans, P. [2009a]. "Global measurement of water waves by Fourier transform profilometry". In: *Experiments In Fluids* 46 (6), pp. 1037–1047.

— [2009b]. "Global measurement of water waves by Fourier transform profilometry". In: *Experiments in Fluids* 46 (6), pp. 1037–1047.

Cox, C. S. [1958]. "Measurement of slopes of high-frequency wind waves". In: *Journal of Marine Research* 16, 199225.

Cox, C. and Munk, W. [1954]. "Measurement of the roughness of the sea surface from photographs of the suns glitter". In: *JOSA Journal of the Optical Society of America* 44 (11), pp. 838–850.

Dermanis, A. [1994]. "Free network solutions with the DLT method." In: *ISPRS Journal of Photogrammetry and Remote Sensing* 49 (4), pp. 2–12.

Deseilligny-Pierrot, M. and Clery, I. [2011]. "Apero, an open source bundle adjustment software for automatic calibration and orientation of set of images". In: *ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (5), pp. 269–276.

Detchev, I., Habib, A., and El-Badry, M. [2013]. "Dynamic beam deformation measurements with off-the-shelf digital cameras". In: *Journal of Applied Geodesy* 7 (3), pp. 147–157.

Dorrington, A. A., Jones, T. W., Danehy, P. M., Watson, K. A., Connell, J. W., Pappa, R. S., and Belvin, K. W. [2010]. "Photogrammetric system and method used in the characterization of a structure". Pat. 7667847.

Ebner, H. [1976]. "Self-calibrating block adjustment." In: *Bildmessung und Luftbildwesen* 44 (4), pp. 128–139.

El-Hakim, S., Gonzo, L., Voltolini, F., Girardi, S., Rizzi, A., Remondino, F., and Whiting, E. [2007]. "Detailed 3D modelling of castles". In: *International Journal of Architectural Computing* 5 (2), pp. 200–220.

Eren, G., Aubreton, O., Meriaudeau, F., Sanchez Secades, L.-A., Fofi, D., Truchetet, F., and Er, A. [2009]. "Scanning From Heating: 3D Shape Estimation of Transparent Objects from Local Surface Heating". In: *Optics Express* 17 (14), pp. 11457–11468.

Faltinsen, O [1993]. *Sea loads on ships and offshore structures*. Vol. 1. Cambridge university press.

Fischler, M. A. and Bolles, R. C. [1981]. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography." In: *Communications of the ACM* 24 (6), pp. 381–395.

Förstner, W. [1999]. *Determining the interior and exterior Orientation of a single image.* Tech. rep. Institute of Photogrammetry, University Bonn.

Förstner, W. and Wrobel, B. [2006]. "Manual of Photogrammetry". In: ed. by J. C. McGlone. 5th Edition. American Society for Photogrammetry and Remote Sensing. Chap. Mathematical concepts in photogrammetry. Pp. 15–181.

Förstner, W. [2013]. "Graphical Models in Geodesy and Photogrammetry". In: *Photogrammetrie, Fernerkundung, Geoinformation* 4, pp. 255–262.

Fraser, C. S. [1997]. "Digital camera self-calibration". In: *ISPRS Journal of Photogrammetry and Remote sensing* 52 (4), pp. 149–159.

— [2013]. "Automatic camera calibration in close range photogrammetry". In: *Photogrammetric Engineering & Remote Sensing* 79 (4), pp. 381–388.

Fraser, C. [1987]. "Limiting error propagation in network design". In: *Photogrammetric Engineering & Remote Sensing* 53 (5), pp. 487–493.

Glassner, A. S. [1995]. *Principles of digital image synthesis*. Vol. 1. Elsevier.

Godding R.; Hentschel, B. K. K. [2003]. "Videometrie im wasserbaulichen Versuchswesen". In: *Wasserwirtschaft, Wassertechnik* 4.

Grün, A. [1978]. "Progress in photogrammetric point determination by compensation of systematic errors and detection of gross errors". In: *ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 22 (Part 3), pp. 113–140.

Grunert, J. A. [1841]. "Das pothenotische Problem in erweiterter Gestalt nebst über seine Anwendungen in der Geodäsie." In: *Grunerts Archiv für Mathematik und Physik* 1, pp. 238–248.

Halstead, M. A., Barsky, B. A., Klein, S. A., and Mandell, R. B. [1996]. "Reconstructing curved surfaces from specular reflection patterns using spline surface fitting of normals". In: *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, pp. 335–342.

Han, B. and Endreny, T. [2014]. "River Surface Water Topography Mapping at Sub-Millimeter Resolution and Precision With Close Range Photogrammetry: Laboratory Scale Application". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (2).

Hartley, R. and Zisserman, A. [2003]. *Multiple view geometry in computer vision*. Cambridge university press.

Harvey, E. and Shortis, M. [1995]. "A system for stereo-video measurement of sub-tidal organisms". In: *Marine Technology Society Journal* 29 (4), pp. 10–22.

Healey, G. and Binford, T. O. [1988]. "Local shape from specularity." In: *Computer Vision, Graphics, and Image Processing* 42 (1), pp. 62–86.

Hertzmann, A. and Seitz, S. M. [2005]. "Example-based photometric stereo: Shape reconstruction with general, varying brdfs". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (8), pp. 1254–1264.

Hilsenstein, V. [2004]. "Design and implementation of a passive stereo-infrared imaging system for the surface reconstruction of Water Waves". PhD thesis. Ruperto-Carola University of Heidelberg, Germany.

Hinz, S, Stephani, M, Schiemann, L, and Zeller, K [2009]. "An image engineering system for the inspection of transparent construction materials". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 64 (3), pp. 297–307.

Höhle, J. [1971]. "Reconstruction of underwater objects". In: *Photogrammetric engineering* 37 (9), p. 948.

Horaud, R., Dornaika, F., and Lamiroy, B. [1997]. "Object pose: The link between weak perspective, paraperspective, and full perspective." In: *International Journal of Computer Vision* 22 (2), pp. 173–189.

Horn, B. K. [1977]. "Understanding image intensities." In: *Artificial intelligence* 8 (2), pp. 201–231.

Horn, B. K. and Schunck, B. G. [1981]. "Determining optical flow". In: *Technical Symposium East*. International Society for Optics and Photonics, pp. 319–331.

Huang, T. S. [2001]. *Calibration and orientation of cameras in computer vision*. Springer Verlag.

Hughes, B., Grant, H., and Chappell, R. [1977]. "A fast response surface-wave slope meter and measured wind-wave moments". In: *Deep Sea Research* 24 (12), pp. 1211–1223.

Ihrke, I., Kutulakos, K. N., Lensch, H. P., Magnor, M., and Heidrich, W. [2008]. "State of the art in transparent and specular object reconstruction". In: *EUROGRAPHICS STAR - State of the art report*.

Ikeuchi, K. and Horn, B. K. [1981]. "Numerical shape from shading and occluding boundaries". In: *Artificial intelligence* 17 (1), pp. 141–184.

Ikeuchi, K. [1981]. "Determining Surface Orientation of Specular Surfaces by Using the Photometric Stereo Method". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-3 (6), pp. 661–669.

Jacobsen, K. [1982]. "Attempt at obtaining the best possible accuracy in bundle block adjustment." In: *Photogrammetria* 37 (6), pp. 219–235.

Jähne, B. and Riemer, K. [1990]. "Two-dimensional wave number spectra of small-scale water surface waves". In: *Journal of Geophysical Research: Oceans (19782012)* 95 (C7), pp. 11531–11546.

Jähne, B., Klinke, J., and Waas, S. [1994]. "Imaging of short ocean wind waves: a critical theoretical review". In: *Journal of the Optical Society of America* 11 (8), pp. 2197–2209.

Jones, T. W., Dorrington, A. A., Brittman, P. L., and Danehy, P. M. [2003]. "Laser Induced Fluorescence for Photogrammetric Measurement of Transparent or Reflective Aerospace Structures". In: *49th Annual International Instrumentation Symposium (ISA)*.

Journee, J. and Pinkster, J. [2002]. *Introduction in ship hydro-mechanics*. Delft University of Technology.

Kager, H. [1989]. "A Universal Photogrammetric Adjustment System." In: *Grün/Kahmen. Optical 3-D Measurement Techniques*. Wichmann Verlag, Karlsruhe, pp. 447–455.

Kiefhaber, D. [2010]. "Development of a Reflective Stereo Slope Gauge for the measurement of ocean surface wave slope statistics". MA thesis. Ruperto-Carola University of Heidelberg, Germany.

Kiefhaber, D. [2014]. "Optical measurement of short wind waves - from the laboratory to the field". PhD thesis. Ruperto-Carola University of Heidelberg, Germany.

Killian, K. [1955]. "Über das Rückwärtseinschneiden im Raum." In: *Österreichische Zeitschrift für Vermessungswesen* 4, pp. 97–104.

Kohlschütter, E [1906]. "Stereophotogrammetrische Arbeiten, Wellen-und Küstenaufnahmen". In: *Forschungsreise SMS Planet* 7 (3).

Koschitzki, R, Schacht, G, Schneider, D, Marx, S, and Maas, H.-G. [2011]. "Integration of photogrammetry and acoustic emission analysis for assessing concrete structures during loading tests". In: *Proc. SPIE, Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection*, pp. 80850I–80850I.

Kraus, K. [1997]. *Photogrammetry - Advanced Methods and Applications*. Vol. 2. Germany: Dümmler Verlag.

Kraus, K. [2007]. *Photogrammetry: geometry from images and laser scans*. Walter de Gruyter.

Kutulakos, K. N. and Seitz, S. M. [2000]. "A theory of shape by space carving". In: *International Journal of Computer Vision* 38 (3), pp. 199–218.

Kutulakos, K. N. and Steger, E. [2008]. "A theory of refractive and specular 3d shape by light-path triangulation." In: *International Journal of Computer Vision* 76 (1), pp. 13–29.

Laas, W. [1906]. "Messung der Meereswellen und ihre Bedeutung für den Schiffbau". In: *Jahrbuch der Schiffbautechnischen Gesellschaft*. Springer, pp. 391–407.

Lensch, H. P. A., Kautz, J., Goesele, M., Heidrich, W., and Seidel, H.-P. [Apr. 2003]. "Image-based Reconstruction of Spatial Appearance and Geometric Detail". In: *ACM Trans. Graph.* 22 (2), pp. 234–257.

Lin, S.-Y., Mills, J. P., and Gosling, P. D. [2008]. "Videogrammetric monitoring of as-built membrane roof structures". In: *The Photogrammetric Record* 23 (122), pp. 128–147.

Lippmann, T. and Holman, R. [1992]. "Wave group modulations in cross-shore breaking patterns". In: *Coastal Engineering Proceedings* 1 (23).

Luhmann, T. [2010]. "Close range photogrammetry for industrial applications". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (6), pp. 558–569.

Luhmann, T., Robson, S., Kyle, S., and Harley, I. [2011]. *Close range photogrammetry: Principles, methods and applications*. Whittles.

Ma, W.-C., Hawkins, T., Peers, P., Chabert, C.-F., Weiss, M., and Debevec, P. [2007]. "Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination". In: *Proceedings of the 18th Eurographics conference on Rendering Techniques*. Eurographics Association, pp. 183–194.

Maas, H.-G. [1995]. "New developments in multimedia photogrammetry". In: *Optical 3D measurement techniques III*. Wichmann Verlag.

— [2008]. "Close range photogrammetry sensors." In: *Advances in Photogrammetry, Remote Sensing and Spatial Information Science: 2008 ISPRS Congress Book*. Ed. by Li, Chen, and Baltsavias. CRC Press, pp. 63–72.

Maas, H.-G. and Hampel, U. [2006]. "Photogrammetric techniques in civil engineering material testing and structure monitoring". In: *Photogrammetric Engineering & Remote Sensing* 72 (1), pp. 39–45.

Maas, H.-G., Gruen, A, and Papantoniou, D [1993]. "Particle tracking velocimetry in three-dimensional flows". In: *Experiments in Fluids* 15 (2), pp. 133–146.

Magnevall, M. [2011]. "Simulation and experimental methods for characterization of nonlinear mechanical systems". PhD thesis. Blekinge Institute of Technology, Sweden.

Maresca, J. W. and Seibel, E. [1976]. "Terrestrial photogrammetric measurements of breaking waves and longshore currents in the nearshore zone". In: *Coastal Engineering Proceedings* 1 (15).

Martinec, D. and Pajdla, T. [2007]. "Robust rotation and translation estimation in multi-view reconstruction". In: *IEEE Conference on Computer Vision and Pattern Recognition.* Pp. 1–8.

Maybeck, P. S. [1982]. *Stochastic models, estimation, and control.* Vol. 1. Mathematics in Science and Engineering 141. UK, London: Academic Press Inc.

McGlone, J. C. [2004]. *Manual of Photogrammetry.* USA: American Society for Photogrammetry and Remote Sensing.

Mehrotra, K. and Mahapatra, P. R. [1997]. "A jerk model for tracking highly maneuvering targets". In: *IEEE Transactions on Aerospace and Electronic Systems* 33 (4), pp. 1094–1105.

Menna, F and Troisi, S [2010]. "Low cost reverse engineering techniques for 3D modelling of propellers". In: *ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (Part 5), pp. 452–457.

Menna, F., Nocerino, E., Troisi, S., and Remondino, F. [2013]. "A photogrammetric approach to survey floating and semi-submerged objects". In: *Proc. SPIE, Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection* 8791, 87910H–87910H–15.

Meriaudeau, F., Sanchez Secades, L., Eren, G., Er, A., Truchetet, F., Aubreton, O., and Fofi, D. [2010]. "3D Scanning of Non-Opaque Objects by means of Imaging Emitted Structured Infrared Patterns". In: *IEEE Transactions on Instrumentation and Measurement* 59 (11), pp. 2898–2906.

MicMac, Apero, Pastis and Other Beverages in a Nutshell! accessed: March, 2015. url: logiciels.ign.fr/?Telechargement,20.

Mikhail, E. M., Bethel, J. S., and McGlone, J. C. [2001]. *Introduction to modern photogrammetry.* Vol. 1. John Wiley & Sons Inc.

Mikhail, E. and Ackermann, F. [1976]. *Observations and Least Squares.* New York, USA: Dun Donnelley.

Montes, R. and Ureñ, C. [2012]. *An Overview of BRDF Models.* Tech. rep. Dep. for Computer Languages and Systems, University of Granada.

Morris, N. J. and Kutulakos, K. N. [2011]. "Dynamic Refraction Stereo". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (8), pp. 1518–1531.

Morris, N. J. [2011]. "Shape Estimation under General Reflectance and Transparency". PhD thesis. University of Toronto.

Mostofi, N, Samadzadegan, F, Roohy, S., and Nozari, M [2012]. "Using vision metrology system for quality control in automotive industries". In: *ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 39, pp. 33–37.

Moulon, P., Monasse, P., and Marlet, R. [2013]. "Global fusion of relative motions for robust, accurate and scalable structure from motion". In: *IEEE International Conference on Computer Vision (ICCV)*. IEEE, pp. 3248–3255.

Mulsow, C [2010]. "A flexible multi-media bundle approach". In: *ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38 (5), pp. 472–477.

Mulsow, C., Maas, H.-G., Westfeld, P., and Schulze, M. [2008]. "Triangulation methods for height profile measurements on instationary water surfaces". In: *Journal of Applied Geodesy* 2 (1), pp. 21–29.

Murase, H. [1992]. "Surface Shape Recontruction of a Nonrigid Transparent Object Using Refraction and Motion". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (10), pp. 1045–1052.

Nayar, S. K., Ikeuchi, K., and Kanade, T. [1991]. "Surface reflection: physical and geometrical perspectives". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13 (7), pp. 611–634.

Nex, F. and Remondino, F. [2014]. "UAV for 3D mapping applications: a review". In: *Applied Geomatics* 6 (1), pp. 1–15.

Nicodemus, F. E., Richmond, J. C., Hsia, J. J., Ginsberg, I. W., and Limperis, T [1977]. *Geometrical considerations and nomenclature for reflectance*. Tech. rep. 160. National Bureau of Standards.

Nistér, D. [2003]. "An efficient solution to the five-point relative pose problem". In: *Proceedings of Computer Vision and Pattern Recognition*. Vol. 2, pp. 195–202.

Nocedal, J. and Wright, S. [2006]. *Numerical optimization, series in operations research and financial engineering*. New York, USA: Springer.

Nocerino, E., Ackermann, S., Del Pizzo, S., Menna, F., and Troisi, S. [2011]. "Low-cost human motion capture system for postural analysis onboard ships". In: *Proc. SPIE, Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection*, pp. 80850L–80850L.

Oren, M. and Nayar, S. K. [1997]. "A theory of specular surface geometry". In: *International Journal of Computer Vision* 24 (2), pp. 105–124.

Otepka, J. [2004]. "Precision Target Mensuration in Vision Metrology." PhD thesis. Technische Universtiät Wien.

Pappa, R. S., Jones, T. W., Black, J. T., Walford, A., Robson, S., and Shortis, M. R. [2002]. "Photogrammetry methodology for gossamer spacecraft structures". In: *Sound and Vibration* 36 (8), pp. 12–21.

Pappa, R. S., Jones, T. W., Lunsford, C. B., and Meyer, C. G. [2006]. "In-vacuum photogrammetry of ten meter square solar sail". In: *Experimental Techniques* 30 (3), pp. 46–51.

Park, J. and Kak, A. C. [2004]. "Specularity elimination in range sensing for accurate 3D modeling of specular objects". In: *Proceedings of 2nd International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*, pp. 707–714.

— [2008]. "3D modeling of optically challenging objects". In: *IEEE Transactions on Visualization and Computer Graphics* 14 (2), pp. 246–262.

Piepmeier, J. A. and Waters, J. [2004]. "Analysis of stereo vision-based measurements of laboratory water waves". In: *Geoscience and Remote Sensing Symposium. IGARSS'04.* Vol. 5, pp. 3588–3591.

Quan, L. and Lan, Z. [1999]. "Linear n-point camera pose determination." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (8), pp. 774–780.

Rantoson, R., Stolz, C., Fofi, D., and Mériaudeau, F. [2010]. "3D reconstruction of transparent objects exploiting surface fluorescence caused by UV irradiation". In: *17th IEEE International Conference on Image Processing (ICIP)*, pp. 2965–2968.

Redondo, J., Rodriguez, A, Bahia, E, Falques, A, Gracia, V, Sánchez-Arcilla, A, and Stive, M. [1994]. "Image analysis of surf zone hydrodynamics". In: *Coastal dynamics' 94; proceedings of an International Conference on the Role of the Large Scale Experiments in Coastal Research*.

Remondino, F. and Campana, S., eds. [2014]. *3D Recording and Modelling in Archaeology and Cultural Heritage - Theory and Best Practices*. Archaeopress BAR Publication.

Remondino, F. [2011]. "Heritage recording and 3D modeling with photogrammetry and 3D scanning". In: *Remote Sensing* 3 (6), pp. 1104–1138.

Ressl, C. [2003]. "Geometry, Constraints and Computation of the Trifocal Tensor". PhD thesis. Institut für Photogrammetrie und Fernerkundung. Technische Universiẗ Wien.

Richter-Gebert, J. [2011]. *Perspectives on projective geometry: a guided tour through real and complex geometry.* Springer.

Rocholz, R. [2008]. "Spatiotemporal measurement of shortwind-driven water waves". PhD thesis. Ruperto-Carola University of Heidelberg, Germany.

Roth, S. and Black, M. J. [2006]. "Specular flow and the recovery of surface structure". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2, pp. 1869–1876.

Rupnik, E. and Jansa, J. [2013]. "Experiences in determination of non-rigid body motion in industrial environment using low-cost photogrammetry". In: *Proc. SPIE, Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection* 8791, pp. 87910I–87910I–11.

— [2014]. "Off-the-shelf videogrammetry - a success story". In: *ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 43 (1), pp. 99–105.

Rusu, R. B. [2009]. "Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments". PhD thesis. Computer Science department, Technische Universitaet Muenchen, Germany.

Sanderson, A. C., Weiss, L. E., and Nayar, S. K. [1988]. "Structured highlight inspection of specular surfaces." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10 (1), pp. 44–55.

Sankaranarayanan, A. C., Veeraraghavan, A., Tuzel, O., and Agrawal, A. [2010]. "Specular surface reconstruction from sparse reflection correspondences". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 1245–1252.

Santel, F. [2006]. "Automatische Bestimmung von Wasseroberflächen in der Brandungszone aus Bildsequenzen mittels digitaler Bildzuordnung". PhD thesis. Fachrichtung Geodäsie und Geoinformatik, Univ. Hannover.

Savarese, S. and Perona, P. [2001]. "Local analysis for 3d reconstruction of specular surfaces." In: *Computer Vision and Pattern Recognition* 2, pp. 738–745.

Schneider, J., Läbe, T., and Förstner, W. [2013]. "Incremental real-time bundle adjustment for multi-camera systems with points at infinity". In: *ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 1, W2.

Schooley, A. H. [1954]. "A simple optical method for measuring the statistical distribution of water surface slopes". In: *JOSA Journal of the Optical Society of America* 44 (1), pp. 37–40.

Schulz, H. [1994]. "Retrieving Shape Information from Multiple Images of a Specular Surface". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16 (2), pp. 195–201.

Shafer, S. A. [1985]. "Using color to separate reflection components". In: *Color Research & Application* 10 (4), pp. 210–218.

Shortis, R., M., Ravanbakskh, M., Shaifat, F., Harvey, S., E., Mian, A., Seager, W., J., Edgington, and R., D. [2013]. "A review of techniques for the identification and measurement of fish in underwater stereo-video image sequences". In: *Proc. SPIE, Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection*, 87910G–87910G.

Shortis, M. R., Clarke, T. A., and Robson, S. [1995]. "Practical testing of the precision and accuracy of target image centering algorithms". In: *Photonics East'95*, pp. 65–76.

Shortis, M. and Johnston, G. [1996]. "Photogrammetry: an available surface characterization tool for solar concentrators, Part I: Measurements of surfaces". In: *Journal of Solar Energy Engineering* 118 (3), pp. 146–150.

Snavely, N., Seitz, S. M., and Szeliski, R. [2008]. "Modeling the world from internet photo collections". In: *International Journal of Computer Vision* 80 (2), pp. 189–210.

Stilwell, D. [1969]. "Directional energy spectra of the sea from photographs". In: *Journal of Geophysical Research* 74 (8), pp. 1974–1986.

Stojic, M, Chandler, J., Ashmore, P., and Luce, J [1998]. "The assessment of sediment transport rates by automated digital photogrammetry". In: *Photogrammetric Engineering & Remote Sensing* 64 (5), pp. 387–395.

Stolz, C., Ferraton, M., and Meriaudeau, F. [2012]. "Shape from polarization: a method for solving zenithal angle ambiguity". In: *Optics Letters* 37 (20), pp. 4218–4220.

Sturm, P. and Triggs, B. [1996]. "A factorization based algorithm for multi-image projective structure and motion". In: *Computer Vision – ECCV'96*. Springer, pp. 709–720.

Sturtevant, B [1966]. "Optical depth gauge for laboratory studies of water waves". In: *Review of Scientific Instruments* 37 (1460).

Sutherland, I. E. [1974]. "Three-dimensional data input by tablet". In: *Proceedings of the IEEE*. Vol. 62. 4, pp. 453–461.

Szeliski, R. [2010]. *Computer vision: algorithms and applications*. Springer.

Tang, R. [2013]. "Mathematical methods for camera self-calibration in photogrammetry and computer vision". PhD thesis. University of Stuttgart.

Tang, R., Fritsch, D., and Cramer, M. [2012]. "New rigorous and flexible Fourier self-calibration models for airborne camera calibration". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 71, pp. 76–85.

Tewinkel, G. C. [1963]. "Water depths from aerial photographs". In: *Photogrammetric Engineering* 29 (6), pp. 1421–1428.

Torlegard, K. [1967]. "On the determination of interior orientation of close-up cameras under operational conditions using three-dimensional test objects." PhD thesis. Kungl. Tekniska Höskolan.

Triggs, B., McLauchlan, P. F., Hartley, R. I., and Fitzgibbon, A. W. [2000]. "Bundle adjustment - a modern synthesis." In: Vision algorithms: theory and practice. Springer Berlin Heidelberg, pp. 298–372.

Tsatsomeros, M. J. [2000]. "Principal pivot transforms: properties and applications". In: *Linear Algebra and its Applications* 307.

Van Den Heuvel, F. A., Verwaal, R., and Beers, B. [2006]. "Calibration of fisheye camera systems and the reduction of chromatic aberration". In: *ISPRS International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36.

Veach, E. [1997]. "Robust Monte Carlo methods for light transport simulation." PhD thesis. Stanford University.

Waas, S. [1988]. "Entwicklung eines Verfahrens zur Messung kombinierter Hhen- und Neigungsverteilungen von Wasseroberflenwellen mit Stereoaufnahmen". MA thesis. Ruperto-Carola University of Heidelberg, Germany.

Wang, J. and Dana, K. J. [2006]. "Relief texture from specularities". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (3), pp. 446–457.

Wells Jeffrey M., T. W. J. and Danehy, P. M. [2005]. "Polarization and Color Filtering Applied to Enhance Photogrammetric Measurements of Reflective Surface". In: *Structures, Structural Dynamics and Materials Conference*.

Weyrich, T., Lawrence, J., Lensch, H., Rusinkiewicz, S., and Zickler, T. [2008]. "Principles of appearance acquisition and representation". In: *Foundations and Trends in Computer Graphics and Vision* 4 (2), pp. 75–191.

Wiora, G., Babrou, P., and Männer, R. [2004]. "Real time high speed measurement of photogrammetric targets". In: *Pattern Recognition*. Springer, pp. 562–569.

Wolff, L. B., Shafer, S. A., and Healey, G. E. [1992]. *Shape Recovery Physics-Based Vision - Principles and Practice.* Jones and Bartlett.

Woodham, R. J. [1980]. "Photometric method for determining surface orientation from multiple images". In: *Optical engineering* 19 (1), pp. 191139–191139.

Wrobel, B. P. [2001]. "Minimum solutions for orientation". In: *Calibration and Orientation of Cameras in Computer Vision*. Springer, pp. 7–62.

Zhang, R., Tsai, P. S., Cryer, J. E., and Shah, M. [1999]. "Shape-from-shading: a survey." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (8), pp. 690–706.

Zhang, X. and Cox, C. S. [1994]. "Measuring the two-dimensional structure of a wavy water surface optically: a surface gradient detector". In: *Experiments in Fluids* 17 (4), pp. 225–237.

Zhang, Z. [2000]. "A flexible new technique for camera calibration". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (11), pp. 1330–1334.

Zisserman, A., Giblin, P., and Blake, A. [1989]. "The information available to a moving observer from specularities." In: *Image and vision computing* 7 (1), pp. 38–42.

WWW References

Breeze Systems. accessed: April 2015. url: www.breezesys.com/MultiCamera.

Fusiello, A. [June 2014]. *Structure from motion: state of the art, examples and open issues - the Computer Vision perspective*. Keynote talk given at the ISPRS TC V Symposium. Riva del Garda. url: https://prezi.com/t4kibkvjekh6/isprs-tcv-2014-keynote/.

James P. O'Shea. accessed: August 2014. url: http://vis.berkeley.edu/~joshea/projects/shapefromshading/.

Open Source Computer Vision Library (OpenCV). accessed: November 2014. url: http://opencv.org/.

Point Cloud Library (PCL). accessed: November 2014. url: http://pointclouds.org.

QUALISYS Motion Capture Systems. accessed: November 2014. url: http://www.qualisys.com/.

Vienna Model Basin Ltd. accessed: November 2014. url: http://www.sva.at/.

Ewelina Rupnik

11 rue Hildegard von Bingen
94120 Paris
France

E-mail: erupnik [at] gmail.com
Mobile: +33 640 15 67 65
Born: October 5. 1986

Professional Experience

7/2015–today	**Research Engineer** École Nationale des Sciences Géographiques (ENSG Géomatique) & Institut de Physique du Globe de Paris (IPGP), Paris, France
Focus	Orientation of high-resolution satellite images
11/2013–6/2015	Research Assistant 3D Optical Metrology (3DOM) Bruno-Kessler Foundation (FBK), Trento, Italy Semantic segmentation of high-resolution aerial images for change detection & Airborne oblique multi-camera systems sensor characterization
11/2010–10/2013	Research Assistant Photogrammetry Research Group, Department of Geodesy and Geoinformation (GEO) Vienna University of Technology Close range photogrammetry – orientation, time sequence analyses, diffuse and specular surface modelling
6–9/2009	Project assistant Institute of Photogrammetry and Remote Sensing AGH - University of Science and Technology, Krakow, Poland Analogue cartographic data conversion to GIS databases
7–9/2007	IAESTE traineeship State Geodetic Administration Office, Zagreb & Rijeka, Croatia Image georeferencing, analogue maps scanning, and GPS field measurements

Education

11/2010–9/2015	**PhD in Photogrammetry at Doctoral College ENSYS2030** Vienna University of Technology Diss – Photogrammetric 3D shape reconstruction of diffuse and specular surfaces in time
2005–2010	MSc in Geodesy spec. Photogrammetry and Remote Sensing AGH - University of Science and Technology, Krakow, Poland Dipl – Shape representation and description for pattern recognition in aerial images
2010	Erasmus Programme Department of Photogrammetry and Remote Sensing Munich University of Technology